Robustly Optimal Income Taxation[*]

Maren Vairo[†]

January 25, 2025

Abstract

We study the design of a tax rule to optimally redistribute income across heterogeneous workers. Traditional models of optimal taxation assume that the social planner has precise knowledge of workers' productivity and preferences, yet such information is often difficult to obtain in practice. We relax this assumption by developing a framework in which the planner faces unquantifiable uncertainty about workers' income opportunities, modeled as ambiguity regarding the economy's production technology. Additionally, we account for individual-level income risk, which introduces a moral hazard component to the planner's problem. Using a worst-case (max-min) criterion, we show that a progressive tax rule—with weakly increasing marginal tax rates—is optimal, regardless of the planner's redistribution preferences or assumptions about the distribution of workers' productivity. Under a richness condition on income opportunities, we further establish that progressivity is necessary for optimality. Unlike the canonical Mirrlees model, our approach yields a strictly positive top marginal tax rate and an everywhere progressive tax schedule. Beyond taxation, our insights contribute to the study of robust mechanism design in settings involving both adverse selection and moral hazard.

[*] I am deeply grateful to Eddie Dekel and Bruno Strulovici for their constant guidance and support, as well as to Piotr Dworczak, Yingni Guo, Asher Wolinsky, and Alessandro Pavan for many detailed comments and discussions at different stages of this project. I thank Ian Ball, Marcus Berliant, Benjamin Brooks, Modibo Camara, Gabriel Carroll, Simone Cerreia-Vioglio, Hector Chade, Joyee Deb, Marina Halac, Ian Jewitt, Peter Klibanoff, Elliot Lipnowski, Alessandro Lizzeri, Wojciech Olszewski, Francisco Pareschi, Harry Pei, Alex Rees-Jones, Ludvig Sinander, Juuso Toikka, Udayan Vaidya, and Weijie Zhong, as well as several seminar audiences, for helpful comments and suggestions.

[†] Deparment of Economics, Princeton University. Email: mvairo@princeton.edu

1 Introduction

Labor income taxation is a crucial tool for governments to redistribute income and provide workers with social insurance against income shocks. Consistent with their relevance, reforms of the income tax system are constantly under debate in most OECD economies (Piketty and Saez, 2013). At the center of this debate is the so-called equity-efficiency tradeoff: a more progressive income tax may be beneficial for social welfare as it contributes to a more equitable distribution of well-being across the population; however, it may also negatively affect efficiency by distorting people's incentives to work. An extensive theoretical literature, dating back to Ramsey (1927) and Mirrlees (1971), has been devoted to the optimal resolution of this tradeoff.

For the government, designing the income tax is an inherently *informationally demanding* task. It requires detailed knowledge of workers' preferences over consumption and leisure, as well as their productivity—both of which are highly idiosyncratic. The existing optimal taxation literature assumes that the social planner has perfect aggregate-level knowledge of these details. However, as Mirrlees (1971) emphasized, such information is "difficult to estimate in real economies." For example, tax schedules are typically rigid, limiting governments' ability to experiment with the tax code, which in turn makes it difficult to infer labor supply elasticities without strong parametric assumptions. Moreover, economic primitives such as preferences and technology evolve over time, reducing the usefulness of historical income data for informing current policy decisions.

In this paper, we relax the assumption that the social planner has complete knowledge of the economic primitives and study the design of optimal income taxes when the planner faces uncertainty about these fundamentals. Specifically, we consider a social planner tasked with designing a labor income tax to maximize a weighted sum of aggregate workers' welfare and tax revenue. As in Mirrlees (1971), workers in the economy are heterogeneous in their ability to generate income and their relative preferences over consumption and leisure. In our framework, a worker's productivity determines the set of *income opportunities* available to them—i.e., the set of output levels they can produce and the corresponding labor disutility costs. Workers' preferences are assumed to be quasilinear in consumption and separable in leisure and consumption. If the planner does not face any uncertainty regarding the primitives of the environment, the Mirrlees framework—or, more precisely, the version analyzed by Diamond (1998b), where preferences are quasilinear—is a special case of the economy just described. In that setting, the only information constraint faced by the social planner stems from the fact that workers' productivity is unobservable.

Our primary departure from this framework is that, in our model, the social planner faces *unquantifiable uncertainty* regarding the set of income opportunities available to each type of worker, a concept we refer to as the economy's *production technology*. The social planner is assumed to have only partial knowledge of this production technology: she knows a minimal set of income opportunities, referred to as the *baseline technology*, that are available to each type of worker but cannot rule out the possibility that workers may choose opportunities outside this set. Accordingly, the social planner considers all possible production technologies that satisfy this condition and designs a tax rule to maximize worst-case welfare across these scenarios.

A second point of departure from Mirrlees (1971) stems from the possibility that workers face individual-level income risk. In many real-life occupations, income is subject to some degree of randomness due to factors beyond the worker's control, such as unforeseen shocks or performance-based pay. Reflecting this reality, the set of production technologies considered by the planner in our model *allows for* (but does not restrict to) the possibility that workers labor effort may lead to *non-degenrate* income lotteries. This feature introduces a moral hazard component into the planner's problem, as taxes can only be conditioned on *realized* income.

To better illustrate the environment, consider a special case of our model in which productivity is determined by a worker's level of education, θ. Suppose that the planner knows the distribution of education levels in the population. However, she faces residual uncertainty about how education influences a worker's income-earning ability—e.g., what labor opportunities are available and the labor disutility of income production. Consequently, the planner cannot fully predict the income choice of a worker with type θ in response to the tax schedule she implements. Such uncertainty could arise from individual-level income shocks (e.g., illness or job loss) or aggregate-level shocks (e.g., an economic recession), as well as from limited knowledge by the government about the underlying primitives. In light of this uncertainty, the planner cannot rely on a parametric relationship between workers' education and their income production, or how it responds to changes in taxation.

Our social planner thus faces three main information constraints. The first arises from the unobservability of workers' productivity: the planner cannot directly condition the income tax on productivity. In our model, as in Mirrlees (1971), the planner holds a well-formed belief about the distribution of this uncertainty and addresses it using a Bayesian criterion. The second constraint concerns the potential for moral hazard: although workers' effort choices may result in random output, taxes can only be conditioned on realized income. The final constraint stems from the planner's limited knowledge of the economy's production technology. We assume that this form of uncertainty is unquantifiable—i.e., the planner does not have a Bayesian belief over the space of production technologies—and the planner addresses it using a max-min criterion.

Our main finding (Theorem 1) is that *irrespective of the planner's preference for redistribution and the distribution of workers' heterogeneity*, the planner can maximize worst-case welfare by implementing a *progressive tax rule*, meaning a tax schedule with weakly increasing marginal tax rates. This contrasts with the canonical Mirrleesian model, where optimal taxes are not generally progressive. For example, in that framework, except in very specific cases (Diamond, 1998b; Saez, 2001), a standard no-distortion-at-the-top argument implies that if workers' productivity is bounded above, the optimal tax rule sets the top marginal tax rate at zero. In contrast, in our model, the optimal top marginal tax rate is generally strictly positive.[1] Beyond the taxation of top incomes, we show that *the entire* tax schedule is progressive at the optimum—a result that, to the best of our knowledge, does not emerge in existing versions of the Mirrlees model.

The intuition behind the optimality of tax progressivity can be better understood by distin-

[1] It is zero only in the extreme case where the planner optimally raises revenue solely through lump-sum payments. We provide conditions under which the tax schedule is not flat in Proposition 2.

guishing two key forces at play in the planner's problem, both of which favor a convex tax schedule. The first force arises from workers' (worst-case) incentives to make risky income choices. We show that strict concavity (i.e., non-progressivity) of the tax schedule leads, in the worst case, to workers engaging in income risk-taking that reduces tax revenue without enhancing their utility. The high-level intuition is that a non-progressive tax schedule introduces a wedge between the planner's and the workers' preferences over income risk. Specifically, concavity of the tax schedule results in a convex after-tax income for workers, making them effectively risk-loving. In contrast, the planner—whose revenue function under a non-progressive tax is concave—becomes effectively risk-averse, thereby incurring losses from income randomness.

Moreover, given the uncertainty about the production technology, the planner cannot rule out the possibility that these revenue-minimizing random income choices will become available and attractive to workers. Indeed, under the worst-case scenario, we show that these are precisely the choices that workers will make. To hedge against such adversarial income risk, the planner can adopt a weakly convex tax schedule, which better aligns the workers' individual incentives with overall welfare. In the absence of worker heterogeneity, we elaborate on this rationale to show the optimality of an affine tax rule, consistent with the findings of Carroll (2015).[2] However, as we describe next, in our model, the introduction of worker heterogeneity requires the planner to consider more complex tax schedules.

Indeed, the second key force driving the optimality of tax progressivity arises from heterogeneity in workers' productivity. To understand this part of the intuition, consider first a benchmark scenario in which workers' productivity is observable and the tax can directly condition on it. As discussed in the previous paragraph, the planner would want to use productivity-specific affine taxes. Moreover, she would likely impose steeper taxes on more productive workers, as the welfare cost of incentivizing them to generate income is lower. Naturally, when workers' types are *unobservable*, adjustments to the tax schedule are necessary to accommodate their private incentives. One way the planner may address adverse selection is by exploiting the fact that more productive workers, in the worst case, choose income levels that result in higher expected income. Consequently, by convexifying the tax schedule—that is, making marginal tax rates *strictly* increasing in income—the planner can indirectly impose higher marginal tax rates on more productive workers. This approach enables her to replicate, at least partially, the steeper taxes she would impose if workers' productivity were observable.

Building on these arguments, Theorem 1 establishes that, independent of the planner's preference for redistribution, any non-progressive tax schedule can be replaced by a progressive one that simultaneously improves both revenue and welfare for every worker. If, in addition, the planner is inequality-averse—meaning her objective places greater marginal value on the welfare of worse-off workers—an additional force favoring progressive taxation emerges. Specifically, the marginal wel-

[2] Carroll (2015) shows that, under a revenue-maximization objective, *linear* taxes—i.e., affine with zero intercept—are optimal. In our setting, the principal's objective is a combination of workers' welfare and revenue, which may justify a positive lump-sum payment to the workers.

fare cost of increasing taxes at the top diminishes when the objective assigns relatively less weight to the utility of higher-income workers. This strengthens the case for progressivity, as it allows the planner to raise more revenue from top earners while incurring a lower welfare loss. This last point highlights that, even though we show that (weak) tax progressivity is optimal under great generality, the actual shape of the optimal tax—e.g., the extent to which marginal tax rates increase with income—is, as one would expect, sensitive to the details of the economy, such as the distribution of workers' types and the government's attitude toward inequality.

Having established that progressivity is a key qualitative property of the optimal tax rule, we proceed to further characterize its precise shape. Deriving closed-form expressions for the optimal marginal tax rate is challenging because workers' income choices are random, and the planner's objective follows a max-min criterion. As a result, the standard optimal control techniques employed by Mirrlees (1971) do not apply in our setting. Instead, we adopt an approach similar in spirit to that of Saez (2001), which involves deriving necessary conditions for optimality by analyzing the effects of small perturbations around a candidate optimal tax rule. A key technical challenge in our model, however, is that—unlike in the canonical framework—the economy's production technology is not fixed. Specifically, a perturbation in the tax schedule affects both (i) the production technology that attains the worst-case scenario and (ii) the workers' income choices under any given production technology. By applying an envelope theorem for arbitrary choice sets (building on Milgrom and Segal (2002)), we establish that the planner's worst-case payoff is (directionally) differentiable with respect to tax perturbations and characterize the resulting derivatives. This result can be adapted to derive necessary conditions for worst-case optimality in other settings involving ambiguity about agents' choice sets.

With this method, we solve for the optimal affine tax and identify conditions under which affine taxes are strictly suboptimal. In that way, we give conditions under which the optimal tax must be *strictly progressive*, in the sense of involving strictly higher marginal tax rates at the top of the income distribution than at the bottom. In deriving these conditions, following the approach of Saez (2001), we distinguish two types of effects of tax perturbations: a *mechanical* effect, reflecting the welfare impact when workers' choices remain fixed, and a *worst-case behavioral* effect, capturing changes in workers' labor decisions due to tax changes. Since the latter focuses on adversarial income choices by workers, its interpretation differs significantly from its Bayesian counterpart. Furthermore, by specializing to the case where the baseline technology corresponds to that of a classical Mirrlees economy with finitely many types, we provide a characterization of the optimal marginal tax rates.

Overall, this paper contributes to the literature on optimal labor income taxation by incorporating the realistic constraint that the government's information about workers is inherently limited. Consistent with real-life practices, we show that the optimal tax schedule under a worst-case approach is progressive, thereby providing a theoretical foundation for progressive taxation in scenarios where the government faces partial uncertainty about economic fundamentals. In this regard, our work joins a growing body of literature showing that intuitive or simple mechanisms can be optimal when the designer operates under environmental uncertainty and evaluates it using a max-min objective (Chung and Ely, 2007; Chassang, 2013; Carroll, 2015).

That said, we do not aim to claim that worst-case concerns are the primary explanation for why progressive taxes are adopted in practice. Instead, we view our framework as a natural benchmark for exploring scenarios where the government lacks detailed information about how workers respond to taxation. The standard Bayesian approach to the taxation problem typically focuses on a specific type of economy—such as the one proposed by Mirrlees (1971)—within which tax progressivity is shown to be suboptimal. In contrast, robustness with respect to the production technology leads us to consider a broader set of environments. In doing so, we identify a class of technologies that, in our view, are both realistic *and* justify the optimality of progressivity. Indeed, our discussion in Section 3.2 illustrates that the case for progressivity extends beyond the specifics of the max-min criterion and instead emerges from broader economic forces that arise when considering a richer set of technologies. There, we describe the structure of the worst-case technology under a non-progressive tax and argue why progressivity may still be beneficial even within a Bayesian framework that departs from the canonical technology.

Beyond its application to income taxation, this paper extends the model of Carroll (2015) by introducing *worker heterogeneity*, thereby offering a framework for studying robust principal-agent contracting under a combination of moral hazard and adverse selection. Accordingly, the model can be applied to other contexts where these features are prevalent, such as the design of managerial compensation schemes, financial securities, and monopoly regulation. In these broader settings, our main result provides a theoretical foundation for the use of concave contracts.

The rest of the paper is structured as follows. In the remainder of this section, we discuss related literature. In Section 2, we describe the model. Section 3 presents our main result on the optimality of progressive taxation, where we also show that, under certain conditions, *any* optimal tax rule must be progressive and provide sufficient conditions for strict progressivity. Section 4 examines a special case of the model in which the baseline technology corresponds to the one assumed in the canonical Mirrlees model. We conclude in Section 5 by discussing extensions of the model and the robustness and limitations of the analysis. All proofs are included in Appendix A.

1.1 Related literature

Mirrlees (1971) introduced the canonical framework for studying labor income taxation, where a utilitarian social planner uses distortionary income taxes to optimally redistribute income among workers with heterogeneous skills. The planner observes and taxes workers' earnings but cannot observe their skills or effort, leading to a non-trivial equity-efficiency trade-off. Mirrlees derived the optimal non-linear tax schedule as a function of workers' earnings. While the solution is complex and sensitive to assumptions about the economy, subsequent work has identified general qualitative features of the optimal tax. Piketty and Saez (2013) provide an extensive review of this literature. Among the most notable findings is the result that, if the distribution of workers' skills is bounded, the marginal tax rate should be zero at the top of the income distribution (Sadka, 1976; Seade, 1977). This (in)famous result has been contested by Diamond (1998b) and Saez (2001), who demonstrated that the zero-taxation-at-the-top prediction does not hold when the right tail of the skill distribution follows a Pareto distribution.

Our main departure from this literature is allowing for unquantifiable uncertainty about the economy's primitives and the use of a max-min criterion to determine the optimal tax.[3] Our assumption of limited information is conceptually related to the *robust control theory* of Hansen and Sargent (2001), which models uncertainty about the data-generating process of shocks. Recent work by Bhandari et al. (2024) applies this approach to taxation, showing that small distributional uncertainty in productivity can *reduce* tax progressivity. Lockwood et al. (2021) and Berliant and Gouveia (2022) incorporate different robustness concerns, studying redistributive taxation under Bayesian uncertainty about economic fundamentals and finite-sample uncertainty about the realization of workers' productivity, respectively.

A second key distinction from Mirrlees (1971) is that we allow for stochastic output, introducing a moral hazard component into the planner's problem. The importance of incorporating income risk into taxation models has long been recognized (Mirrlees, 1974; Varian, 1980).[4] Most prior studies addressing income risk focus on homogeneous workers, framing the problem as one of balancing incentives and insurance provision.[5] In contrast, our framework jointly accommodates moral hazard and adverse selection, both of which represent realistic informational constraints faced by tax authorities. Additionally, relative to the canonical Mirrlees framework, we allow for considerably richer worker heterogeneity: instead of modeling "productivity" as a one-dimensional type, we represent it as a *set* of income opportunities available to each worker.

More broadly, our work relates to mechanism design with redistributional concerns. Recent contributions to this topic include studies on dynamic income taxation (Farhi and Werning, 2013; Golosov et al., 2016; Stantcheva, 2017; Makris and Pavan, 2021), redistribution in two-sided markets (Dworczak et al., 2021), allocation of public resources (Akbarpour et al., 2024; Kang, 2023), and taxation of externalities (Pai and Strack, 2023).

Additionally, we contribute to the literature on robust contracting and mechanism design with a worst-case objective, surveyed by Carroll (2019). We build on the framework of Carroll (2015), who studies robust contracting under moral hazard and ambiguity about the agent's production technology. Our model differs by incorporating both moral hazard and adverse selection, with the latter arising from heterogeneous sets of feasible actions across workers. Furthermore, our principal's objective combines welfare and revenue maximization, unlike the purely revenue-maximizing principal in Carroll (2015). We extend Carroll's result on the optimality of linear contracts by showing that adverse selection justifies the use of *concave contracts*.[6]

[3] This contrasts with studies on taxation with max-min preferences, where the government has complete information about the economy but adopts a Rawlsian welfare criterion (Phelps, 1973; Boadway and Jacquet, 2008).

[4] See Chapter 11 in Tuomala (2016) for a review.

[5] An exception is Boadway and Sato (2014), who show that introducing moral hazard into the Mirrlees (1971) framework with adverse selection may result in negative marginal taxes at the top, leading to an extremely *regressive* optimal tax.

[6] Here, concave contracts correspond to convex (progressive) taxes.

A related point is made by Barron et al. (2020) (see also Diamond (1998a)) in a Bayesian moral hazard model without adverse selection. In those papers, the principal knows the production technology and assumes the agent can garble any output realization at zero cost. This introduces a "no gaming" constraint, requiring the principal to anticipate the agent's gambles by concavifying the offered contract. In our model, worker heterogeneity and richer uncertainty lead the principal to hedge against additional distortions—beyond mean-preserving spreads—that different worker types might use to alter the output distribution.[7]

Although our primary focus is income taxation, our results apply to broader models of robust contracting with moral hazard and adverse selection. For instance, Carroll and Meng (2016) provide a foundation for linear contracts in a model in a model where ambiguity over income shocks interacts with private information held by the agent. The nature of ambiguity in their paper differs considerably from ours, leading to qualitatively different results about the optimal contract. We also relate to the work by Garrett (2014), who studies robust contracting for cost-based procurement. Like ours, that model features a principal who must jointly address worst-case uncertainty and the classic Bayesian trade-off between efficiency and information rents. In Garrett's setting, the optimal mechanism involves fixed-price cost-reimbursement contracts, whereas progressivity achieves the optimal trade-off in our taxation model with uncertain production technology.

2 Model

Preliminaries. In what follows, for a metric space X we use $\Delta(X)$ to denote the set of Borel probability distributions on X, and we equip $\Delta(X)$ with the topology of weak convergence. For any $x \in X$, we use δ_x to denote the Dirac measure on $\{x\}$. The real line is endowed with the Euclidean topology and the product space $\Delta(X) \times \mathbb{R}$ is endowed with the respective product topology. Any space of functions from X to $\mathbb{R}$ is equipped with the sup-norm topology.

2.1 General framework

There is a social planner (she), and a non-atomic unit mass of workers (he, for the singular) indexed by $i \in \mathcal{I} \equiv [0,1]$ uniformly distributed on the unit interval.[8] Each worker i produces income $y_i \in Y \equiv [0,\overline{y}]$, where $\overline{y} \in \mathbb{R}_{++}$. The social planner designs an *income tax rule*, mapping income realizations into payments made by the worker. Formally, a tax rule is a function $T : Y \to \mathbb{R}$. A worker's consumption, as a function of his earned income, equals after-tax income $y - T(y)$. A tax rule T is said to be *feasible* if it is continuous and satisfies $y - T(y) \geq 0$ for all $y \in Y$.[9] In line with

[7] In their Online Appendix, Walton and Carroll (2022) examine a moral hazard setting where the principal's robust concern is limited to the agent's garbling of output. They show that this results in the optimality of a concave contract.

[8] Because an agent's index i does not carry any meaning, this is a normalization of an arbitrary population of workers described by a non-atomic measure space $(G, \mathcal{G}, \gamma)$ where G has the cardinality of $\mathbb{R}$.

[9] We restrict to continuous tax schedules to simplify exposition: With some extra verifications, our results can be extended to the case in which only lower semi-continuity of T is required.

the bulk of the optimal taxation literature, we restrict attention to deterministic tax rules.[10] We let $\mathbf{T}_f$ denote the set of feasible tax rules.

Workers' labor decision. Conditional on the tax rule, workers make labor decisions to maximize their utility. Specifically, a worker's income choice consists of a pair (F, ϕ) where $F \in \Delta(Y)$ is the worker's stochastic earned income and $\phi \in \mathbb{R}_+$ is the labor disutility associated with the income lottery F.[11]

The economy's *production technology* determines the income choices that are available to every worker. Specifically, let $\mathcal{M}$ be the set of non-empty, compact subsets of $\Delta(Y) \times \mathbb{R}_+$. Each worker $i \in \mathcal{I}$ is endowed with a set of income choices $M_i \in \mathcal{M}$ that he can choose from. The economy's *production technology* is a collection of workers' feasible sets $\mathbf{M} = \{M_i\}_{i \in \mathcal{I}} \in \mathcal{M}^{\mathcal{I}}$.

Workers' preferences are assumed to be quasilinear in consumption, and additively separable in consumption and labor. Therefore, given the tax rule T and the technology $\mathbf{M}$, the income choice of worker i is described by

$$V_w(T|M_i) \equiv \max_{(F,\phi)\in M_i} \{\mathbb{E}_F[y - T(y)] - \phi\}, \quad X_w(T|M_i) \equiv \operatorname*{arg\,max}_{(F,\phi)\in M_i} \{\mathbb{E}_F[y - T(y)] - \phi\}.$$

The following example illustrates that the canonical model by Mirrlees (1971) can be described using our terminology and therefore is a special case of the environment described so far.

Example 1 (Mirrlees economy)**.** Suppose that $\theta_i \in \Theta \subset \mathbb{R}_{++}$ represents the wage of worker i. Let $\mu \in \Delta(\Theta)$ be the population-distribution of θ_i and let $l_i \in [0,1]$ denote the units of labor provided by worker i. Given his wage θ_i and his labor choice l_i, worker i's earned income is assumed to be *deterministic* and equal to $y_i = l_i \times \theta_i$. There is a continuous, increasing function $\Phi : [0,1] \to \mathbb{R}_+$ that governs workers' labor disutility. This setting can be embedded into ours by writing the production technology as

$$M_i = \{(\delta_y, \phi) \in \Delta(Y) \times \mathbb{R}_+ : y \le \theta_i, \phi = \Phi(y/\theta_i)\}.$$

Social planner's information. Our main point of departure from Mirrlees (1971) is our assumption that the social planner faces uncertainty about the production technology. We model this uncertainty by assuming that she only knows a *baseline technology* $\mathbf{M}^0 = \{M_i^0\}_{i\in\mathcal{I}} \in \mathcal{M}^{\mathcal{I}}$ which describes a minimal set of actions available to each worker. Formally, we assume the social planner entertains any production technology $\mathbf{M}$ belonging to the *plausible set* as defined in Definition 1.

Definition 1 (Plausible technologies)**.** The set of plausible technologies $\mathcal{M}_p$ consists of all $\{M_i\}_{i\in\mathcal{I}} \in$

[10] As shown by Kambhampati (2023) and Kambhampati et al. (2024), this restriction will typically entail a loss of optimality in the type of setting that we study.

[11] This framework is consistent with a more standard interpretation in which workers choose a level of effort e which leads to a pair $(F(e), \phi(e)) \in \Delta(Y) \times \mathbb{R}_+$. To save on notation, we do not explicitly model effort and instead assume that workers directly choose a pair (F, ϕ) from their feasible set described below.

$\mathcal{M}^{\mathcal{I}}$ such that: *(i)* the correspondence $i \rightrightarrows M_i$ is weakly measurable,[12] and *(ii)* for all $i \in \mathcal{I}$, $M_i \supseteq M_i^0$.

The first condition in Definition 1 is a mild technical requirement to ensure that the social planner's objective is well-defined. The second part of the condition states that there is a known minimal set of income choices that is available to every type of worker which is described by the baseline technology. We make the following assumption on the baseline technology, ensuring that $\mathbf{M}^0$ is itself plausible.

Assumption 1 (Baseline technology)**.** The correspondence $i \rightrightarrows M_i^0$ is weakly measurable.

2.2 Social planner's problem

We now turn to describing the social planner's problem of finding an optimal tax rule. Her objective is a weighted sum of average workers' welfare and tax revenue. Specifically, for a given plausible production technology $\mathbf{M} = \{M_i\}_{i\in\mathcal{I}} \in \mathcal{M}_p$ and a feasible tax rule $T \in \mathbf{T}_f$, let $(F_i, \phi_i) \in M_i$ be the equilibrium income choice of worker i. The planner values worker i's welfare according to the welfare function $W(\mathbb{E}_{F_i}[y_i - T(y_i)] - \phi_i, i)$. Throughout, we assume that $u \to W(u, i)$ is non-decreasing and differentiable for every $i \in \mathcal{I}$, and that $i \to W(u, i)$ is measurable and bounded for every $u \in \mathbb{R}$. The second component in the social planner's objective is expected revenue, which we assume for simplicity that she values linearly with a constant marginal value that we denote by $\alpha \in \mathbb{R}_{++}$. We discuss the role of this assumption and provide an interpretation for α as the government's marginal value of public funds in the following subsection. We assume that, when indifferent, workers break ties in favor of the income choice that maximizes tax revenue.[13]

As a result, the planner's payoff when she offers the tax rule $T \in \mathbf{T}_f$ and the production technology is $\mathbf{M} \in \mathcal{M}_p$ is given by[14]

$$V_P(T|\mathbf{M}) \equiv \int_{\mathcal{I}} \left\{ W(V_w(T|M_i), i) + \alpha \max_{(F_i,\phi_i)\in X_w(T|M_i)} \mathbb{E}_{F_i}[T(y)] \right\} di. \tag{2.1}$$

We refer to $V_P(T|\mathbf{M})$ as *total welfare*, to its first component as *total worker-welfare*, and to its second component as *total revenue*; and oftentimes omit the word "total" for conciseness.

The planner's problem is to choose a feasible tax rule that maximizes her worst-case payoff across all plausible $\mathbf{M}$'s. Formally, the planner's problem is

$$\sup_{T\in\mathbf{T}_f} \inf_{\mathbf{M}\in\mathcal{M}_p} V_P(T|\mathbf{M}). \tag{2.2}$$

[12] To define weak measurability of M_i, let $A \subseteq \Delta(Y) \times \mathbb{R}_+$ and $M^l(A) \equiv \{i \in \mathcal{I} : M_i \cap A \neq \emptyset\}$ be the lower inverse of M_i. We say that $i \rightrightarrows M_i$ is weakly measurable if $M^l(A)$ is Lebesgue-measurable for every open $A \subseteq \Delta(Y) \times \mathbb{R}_+$.

[13] This assumption ensures that the maximum in the planner's problem is attained. If the workers were instead to break ties adversarially, the supremum in the planner's problem is the same, but is not necessarily attained by some tax rule.

[14] Weak measurability ensures that the integral in (2.1) is well-defined for all $M \in \mathcal{M}_p$ and all feasible T by the Measurable Maximum Theorem (Theorem 17.18 in Aliprantis and Border (2006)).

We refer to a tax rule as being *worst-case optimal* if it is feasible and attains the value in (2.2).

To ensure the existence of a worst-case optimal tax, we make the following assumption regarding the relative weights of workers' welfare and revenue in the planner's objective. Let $W_1(u,i)$ denote the partial derivative of $W(u,i)$ with respect to u.

Assumption 2 (α-bound)**.** There exists $C \in \mathbb{R}_+$ such that for any Lebesgue-measurable function $i \to c_i$ such that $c_i > C$ for all $i \in \mathcal{I}$,[15] it holds that $\int_{\mathcal{I}} W_1(c_i, i)\, di \le \alpha$.

2.3 Discussion of the model

In this section, we provide an interpretation of the model and its assumptions. We resume this discussion in Section 5, where we comment on the robustness of our results.

Income risk. One point of departure of our model relative to the standard Mirrleesian framework is that workers are allowed to make labor choices that lead to stochastic income. An interpretation of this assumption is that workers may choose occupations that intrinsically involve income risk. Examples of such jobs abound in practice: e.g., bonus-based compensation schemes, entrepreneurial activity, or freelance work are a few salient ones. Additional sources of randomness in workers' reported income, not stemming from occupational choices, are given by workers' decisions to financially invest part of their labor income, as well as *income spreading* which is the practice of splitting reported income over different years to reduce the tax burden.[16] Therefore, we view this assumption not as a limitation but as a realistic feature of the model. Importantly, our model does not rule out deterministic income choices by workers: We only assume that the social planner *entertains* the possibility that the worker's choice set *includes* non-degenerate income lotteries.

As mentioned in the discussion of the literature, it has been recognized by Mirrlees (1974) that income risk is an important aspect of workers' labor decisions that the original framework in Mirrlees (1971) does not accommodate. Aside from the technical difficulties stemming from the combination of moral hazard and adverse selection, allowing for the possibility that workers' income is random poses the additional challenge of having to take a stance on, first, what set of income lotteries is available to each worker, and second, on workers' labor disutility as a function of these lotteries. In the absence of a well-founded assumption on preferences and technology over income risk, our robust approach arises as a natural, non-parametric way to study the problem.

Assumptions on workers' payoffs Our assumption that workers' preferences are quasilinear in consumption has two implications: First, workers are risk-neutral, and second, there are no income effects in workers' income choices. Under a Bayesian objective, the optimal income taxation problem with quasilinear preferences was first studied by Atkinson (1995) and Diamond (1998b), and has ever since been commonly used by theoretical and empirical work. As in those papers, this (restrictive) assumption enhances the model's tractability and allows us to make progress in

[15] Throughout, we use the expression "for all $i \in \mathcal{I}$" as a shortcut for "for a set of $i \in \mathcal{I}$ of Lebesgue measure one".

[16] Relatedly, Landier and Plantin (2017) study optimal taxation, considering the possibility of tax evasion, which they model as the introduction of random noise to workers' income.

studying the question at hand. We discuss in more detail what happens when the workers' utility of consumption is allowed to be strictly concave in Section 5.

Information structure. Our modeling of the social planner's uncertainty about the production technology builds on Carroll (2015) by introducing worker-heterogeneity. Within the taxation application, this framework can be interpreted as follows. Workers are heterogeneous with regard to their productivity, which in our model is represented by their feasible set of income choices M_i. The critical piece of information that the planner has about the technology is that there is a minimal set of choices available to each worker, as described by the baseline technology $\mathbf{M}^0$. This baseline set of income choices may vary across workers, and we assume that the planner knows how this heterogeneity is distributed in the population. This assumption is reasonable if we interpret M_i^0 as capturing information about workers' demographic characteristics such as education level and wages. Under this interpretation, our assumption is that the planner knows the distribution of demographic characteristics but does not perfectly know the mapping between workers' characteristics, and how much income they can produce and at what cost.

In addition, the uncertainty faced by the planner is rich: Any technology, so long as it contains the baseline and satisfies a mild measurability condition, is plausible. In Section 5, we argue that our main result continues to hold under an alternative information structure in which the uncertainty about $\mathbf{M}$ is considerably smaller.

Assumptions on the social planner's objective. Our assumptions on how the planner values workers' welfare are standard. As in Saez and Stantcheva (2016), we allow for a general welfare objective that may depend both on workers' utility and their identity. In this setting, inequality aversion by the planner would be captured if, for example, $W(u, i)$ is concave in u. However, we have not imposed such assumption and in fact, our main result (Theorem 1) does not rely on any assumptions about the planner's preference for redistribution.

The assumption that the planner values revenue linearly, on the other hand, differs from the standard Mirrleesian model: There, the social planner faces a budget constraint and revenue does not directly enter her objective. In that setting, society's marginal value for revenue is endogenously described by the shadow value associated with the budget constraint.

Under the max-min approach, it is not immediate how to incorporate a hard budget constraint to the problem because whether or not the constraint holds depends on workers' income choices and thus on the technology $\mathbf{M}$, which the planner does not know at the time of designing the tax code. Instead of using the conservative approach of requiring the budget to balance *for every plausible technology* $\mathbf{M}$, we assume that the government tolerates running a budget deficit and strictly values raising more funds. Allowing the government to run a deficit is, in our opinion, a more realistic assumption. For simplicity, we assume that the government's marginal value for public funds is captured by the constant $\alpha > 0$.[17] This assumption is not crucial and our main result (Theorem 1) continues to hold if the planner values total revenue using any non-decreasing function, so long

[17] The same approach is used, in a different setting, by Akbarpour et al. (2024).

as the appropriate version of Assumption 2 is satisfied.[18]

We may interpret α as the value of allocating an additional dollar of tax revenue to alternative projects. Assumption 2 ensures the problem is well-defined by requiring that the planner's marginal value for revenue is not too low compared to the value of lump-sum payments to the average worker. For instance, if $W(u, i) = \omega(i)u$, the assumption implies $\int_{\mathcal{I}} \omega(i), di \leq \alpha$, meaning the planner prefers to use extra revenue for other projects rather than redistributing it as lump-sum payments.[19] This prevents the planner from optimally giving infinite lump-sum payments, enabling the existence of a worst-case optimal tax rule. If $W(u, i)$ is nonlinear in u, the assumption holds under a standard Inada condition, where the marginal value of giving workers an additional dollar approaches zero as after-tax income becomes arbitrarily high.

Available tax rules. We have simplified the social planner's problem by restricting to tax rules that are deterministic and only depend on realized income. We believe that this restriction is in line with real-life tax systems, but in principle, entails a loss in optimality. In particular, because workers effort is not-contractible (i.e., income is random) and workers have private information about the production technology, the planner could benefit from screening this information by offering a *menu* of tax rules. We explore this possibility in Section 5.

3 Main results

In this section, we state our main result regarding the optimality of progressive taxation (Theorem 1), and provide an intuitive discussion of its proof. In all of our discussions, we use the term *progressive* informally to refer to a tax rule $T(y)$ which is convex. We discuss the implications of Theorem 1 in relation to the literature in Section 3.2. Following this, Proposition 1 establishes a sense in which progressivity is necessary for optimality and Proposition 2 provides conditions under which the optimal tax is "strictly" progressive.

3.1 Optimality of progressive taxation

Theorem 1 establishes that there exists an optimal tax that is progressive, and that features non-decreasing tax payments and non-decreasing consumption. We refer to a tax schedule satisfying the latter property—non-decreasingness of tax payments and consumption—as *double-monotone.* Double monotonicity is also satisfied by the optimal tax that arises in the canonical model of Mirrlees (1971). On the other hand, as we argue in Section 3.2, the convexity property is novel and does not arise in that framework.

Theorem 1 (Progressivity)**.** *There exists a worst-case optimal T such that: $T(y)$ is non-decreasing and convex, and $y - T(y)$ is non-decreasing.*

[18] For example, a hard budget constraint can be incorporated by assuming that the value for revenue is zero if the budget is balanced, and $-\infty$ if it is not. More generally, we could set the planner's value for revenue to capture any non-linear costs of government debt.

[19] In Mirrlees (1971), with no alternative public uses for funds, this condition holds with equality.

The formal proof of the theorem is in Appendix A. In this section, we give an intuitive discussion of the proof by proceeding in two steps. First, we provide a simple characterization of the worst-case payoff of the planner (Lemma 1). Second, building on this characterization, we give a graphical description (Figure 3.1) of the other main part of the argument in the proof of Theorem 1. That part consists of constructing a worst-case welfare-improving convex version of an arbitrary tax function. A final step, relegated to the Appendix, establishes that a worst-case optimal tax exists.

Characterizing the worst case. As a key step towards establishing the main result, Lemma 1 describes worst-case welfare under a given tax rule $T \in \mathbf{T}_f$, which we denote by

$$V_P(T) \equiv \inf_{\mathbf{M} \in \mathcal{M}_p} V_P(T|\mathbf{M}).$$

Lemma 1 (Worst case)**.** *For any feasible T,*

$$V_P(T) = \int_{\mathcal{I}} \{W(V_w(T|M_i^0), i) + \alpha \mathbb{E}_{F_i}[T(y)]\}\, di, \tag{3.1}$$

where $F_i \in \Delta(Y)$ is defined by one of the following two cases:

(i) If $V_w(T|M_i^0) < \max_{y \in Y}\{y - T(y)\}$, then

$$F_i \in \arg\min_{F \in \Delta(Y)} \; \mathbb{E}_F[T(y)], \quad \textit{subject to } \mathbb{E}_F[y - T(y)] \geq V_w(T|M_i^0); \tag{3.2}$$

(ii) If $V_w(T|M_i^0) = \max_{y \in Y}\{y - T(y)\}$, then

$$F_i \in \arg\max_{(F,0) \in X_w(T|M_i^0)} \; \mathbb{E}_F[T(y)]. \tag{3.3}$$

According to the first term in (3.1), given the tax rule T, the equilibrium utility of worker i under the worst-case technology is equal to the lower bound on their equilibrium payoff as given by $V_w(T|M_i^0)$.[20] The intuition for this part of the result is simple: If workers' welfare was strictly higher than the welfare-lower-bound $V_w(T|M_i^0)$, it would be possible to construct a plausible technology that uniformly increases their labor cost $\phi \in \mathbb{R}_+$ for every income lottery in their choice set. This constant shift strictly reduces workers' welfare without affecting their incentives to produce income, and hence overall reduces the planner's objective. Therefore, this lower bound is achieved under a worst-case technology.

The second term in (3.1) describes worst-case revenue by distinguishing two cases. The first case pertains to workers who *cannot* attain maximal consumption at zero cost under T and M_i^0. Loosely speaking, Lemma 1 states that under the worst-case scenario, these workers' income choice (F_i, ϕ_i) minimizes their expected tax payment, with ϕ_i defined so that the worker i is indifferent between

[20] The fact that $M_i \supseteq M_i^0$ for all $i \in \mathcal{I}$ and $\mathbf{M} \in \mathcal{M}_p$ implies that $V_w(T|M_i) \geq V_w(T|M_i^0)$.

choosing (F_i, ϕ_i) and choosing the best element in M_i^0. Intuitively, the choice (F_i, ϕ_i) clearly gives a lower bound on the planner's revenue: Under any plausible technology, worker i's income choice (F, ϕ) satisfies $\mathbb{E}_F[y - T(y)] \geq \mathbb{E}_F[y - T(y)] - \phi \geq V_w(T|M_i^0)$. Thus, worker i's "worst-case income choice" is feasible for the minimization program defined in (3.2). To complete the argument, in the proof, we construct a sequence of plausible technologies under which workers' income choice becomes arbitrarily close to (F_i, ϕ_i) as defined above, and in that way, we show that the identified lower bound on the planner's payoff is attained in the limit by some plausible technology.[21] The second case in Lemma 1 concerns workers whose equilibrium payoff is constant in M_i and equal to the highest possible utility that can be attained in the economy. Given that, when indifferent, workers make the income choice that maximizes tax revenue, the adversarial scenario for these types is attained when they face M_i^0 since this minimizes their ability to break ties favorably.

In summary, the worst-case technology is effectively

$$M_i = M_i^0 \cup \{(F_i, \mathbb{E}_{F_i}[y - T(y)] - V_w(T|M_i^0))\},$$

with F_i defined as in Lemma 1, with a qualification that we formally deal with in the proof which is that, as mentioned above, M_i need not attain the worst-case payoff in (3.1), but as we show there is a perturbed version of it that approximates it arbitrarily well.

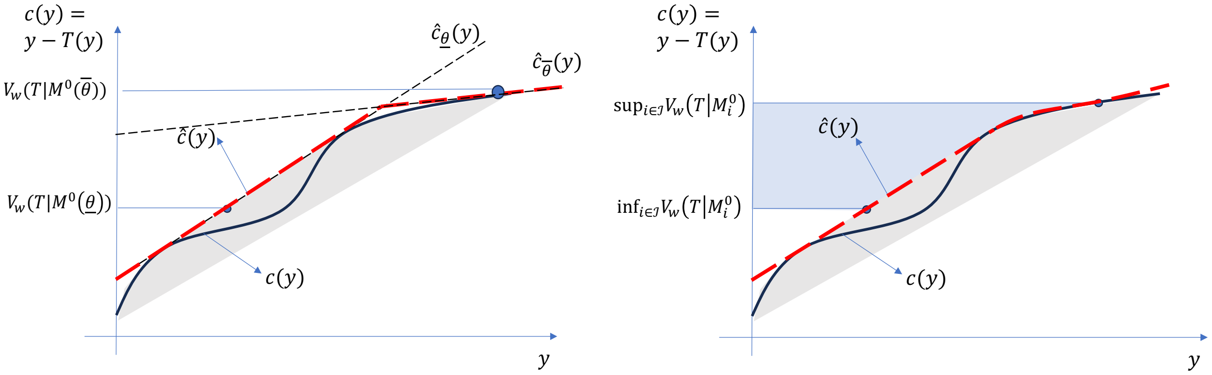
Lemma 1 reduces the objective of the planner in two important ways. First, we have transformed the problem of finding the worst-case payoff under T, which was originally a problem of finding the infimum across plausible correspondences $i \rightrightarrows M_i$, to a much simpler one which consists of finding, for every $i \in \mathcal{I}$, the distribution F_i that satisfies the conditions in the lemma. Second, there are no interactions across workers, and therefore the worst-case scenario can be computed independently for every $i \in \mathcal{I}$. As a result, in the non-trivial case (case (i)), the worst case can be computed by pointwise solving a simple prior-free constrained information design problem.[22] Building on this, in the next step of the intuitive proof, we use simple concavification arguments to graphically depict workers' worst-case income choices.

Construction of a welfare-improving convex tax: A special case. In the proof of Theorem 1, we consider an arbitrary feasible tax T, and use it to construct a convex tax $\overline{T}$ that satisfies $V_P(\overline{T}) \geq V_P(T)$. Here, we informally discuss this construction for the following special case of our framework. Suppose that each worker $i \in \mathcal{I}$ has a binary type $\theta_i \in \Theta \equiv \{\underline{\theta}, \overline{\theta}\}$. A worker's type fully determines his baseline technology. Formally, there exists a correspondence $M^0 : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$ such that $M_i^0 = M^0(\theta_i)$ for all $i \in \mathcal{I}$. Suppose that the social planner uses a tax rule $T(y)$, with associated consumption rule $c(y) = y - T(y)$ as the one depicted in the left-panel of Figure 3.1. The shaded area in the figure consists of all the pairs $(\mathbb{E}_F[y], \mathbb{E}_F[c(y)])$ that can be attained by some income choice $F \in \Delta(Y)$. Formally, it is the convex hull of the graph of $c(y)$.

[21] The reason why the lower bound is not necessarily attained stems from the fact that workers break ties in favor of the social planner. This feature implies that we have to perturb (F, ϕ) so as to ensure that the worker is strictly willing to choose it over choosing an element of M_i^0.

[22] Le Treust and Tomala (2019) and Doval and Skreta (2023) develop techniques for solving this type of problem.

According to Lemma 1, in the worst case, a worker with type $\theta_i \in \Theta$ makes an income choice $F_{\theta_i} \in \Delta(Y)$ that minimizes expected tax revenue subject to the constraint that expected consumption has to be weakly greater than $V_w(T|M^0(\theta_i))$.[23] It follows from well-known arguments (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011) that, for every $\theta \in \Theta$, the revenue minimizing pair $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ is a point in the graph of the upper concave envelope of $c(y)$, which corresponds to the upper boundary of the shaded area in the figure. If, as the example in the figure illustrates, the constraint in the revenue minimization problem binds, then the exact location of $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ is pinned down by the intersection between this upper concave envelope and the horizontal line passing through $V_w(T|M^0(\theta))$.



$c(y) =$
$y - T(y)$
$\hat{c}_{\underline{\theta}}(y)$
$\hat{c}_{\overline{\theta}}(y)$
$V_w(T|M^0(\overline{\theta}))$
$\hat{c}(y)$
$V_w(T|M^0(\underline{\theta}))$
$c(y)$
y
$c(y) =$
$y - T(y)$
$\sup_{i \in \mathcal{I}} V_w(T|M_i^0)$
$\hat{c}(y)$
$\inf_{i \in \mathcal{I}} V_w(T|M_i^0)$
$c(y)$
y

Figure 3.1: An arbitrary consumption rule (solid black line) and a concave consumption rule that dominates it (dashed red line) under a baseline technology with binary types (left-panel), and under an arbitrary baseline technology with a continuum of realizations (right-panel).

Thus, for each $\theta \in \Theta$, $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ belongs to the boundary of the convex hull of the graph of $c(y)$. Building on the arguments in Carroll (2015), the supporting hyperplane to this set passing through the point $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ defines an affine consumption rule that leads to weakly higher revenue *conditional on* $\theta_i = \theta$. These θ-specific consumption rules are depicted by the dashed gray lines in the left-panel of Figure 3.1. In order to construct a consumption rule that outperforms the original one, we seek to aggregate these two affine functions into a single one that outperforms c. We show that this goal is attained by taking the pointwise minimum between the two lines, as given by the red dashed function in the figure. The resulting consumption (tax) function is a two-piece affine function that is concave (convex), and gives weakly higher worst-case revenue than the starting point. Also, as the figure shows, the new consumption rule is pointwise higher than the original one, thereby improving workers' welfare at the same time.

For an arbitrary baseline technology $\mathbf{M}^0$, we apply a similar argument and take the pointwise infimum over the family of i-specific affine consumption rules that support the point $(\mathbb{E}_{F_i}[y], \mathbb{E}_{F_i}[c(y)])$, where F_i is worker i's worst-case income choice, as established in Lemma 1. The right panel of Figure 3.1 illustrates this construction for the case in which minimizing over a continuum of i-specific

[23] In Figure 3.1, the consumption rule is increasing and thus, has a unique global maximizer, which implies that Case (ii) in Lemma 1 does not apply. Our formal proof of Theorem 1 accounts for this case when it arises.

dominating affine rules results in a smooth dominating consumption schedule.

3.2 Discussion and implications of Theorem 1

In this section, we discuss the intuition behind the optimality of tax progressivity and contrast this finding with the canonical optimal taxation model. The intuition can be conveyed through the graphical argument presented in Figure 3.1. The argument highlights two forces that favor the optimality of a convex tax schedule.

The first force arises from workers' incentives to make risky income choices. As shown in the figure, strict concavity of the tax schedule implies that, in the worst case, workers engage in income risk-taking, reducing tax revenue. Intuitively, the concavity of T creates a wedge between the planner's and workers' preferences over income risk: the planner becomes effectively risk-averse (since tax revenue is concave), while workers exhibit risk-loving behavior (because after-tax income is convex). Moreover, whether a worker's income choice is random does not affect the welfare component of the planner's objective. By Lemma 1, this component depends only on the lower bound of workers' payoffs, $V_w(T|M_i^0)$. Thus, the planner should anticipate this adversarial income risk by replacing T with a suitable *weakly* convex version, as constructed in the theorem's proof. In fact, in the absence of worker heterogeneity, a similar intuition to that of Carroll (2015) implies that an affine tax schedule best aligns the workers' and planner's incentives.

The second force arises from heterogeneity in workers' feasible choices, which favors *strict* convexity of the tax schedule. To illustrate this, consider again the special case where workers have types $\theta_i \in \Theta$ and $M_i^0 = M^0(\theta_i)$ for a known correspondence $M^0 : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$. Further, assume that $\Theta \subset \mathbb{R}$ and that M^0 is monotone in the sense of set inclusion. This implies that $V_w(T|M^0(\theta_i))$ is non-decreasing in θ_i. In this context, workers with higher θ_i can be interpreted as having higher *baseline productivity.*

As shown in Figure 3.1, the slope of the dominating type-specific affine tax increases with θ_i. This suggests that if the planner could observe and directly condition taxes on workers' types, it might be optimal to impose steeper taxes on workers with higher θ_i. For further intuition, consider a worker who can achieve any level of income production at zero labor cost under the baseline technology. This worker's baseline utility equals the maximum possible utility, $\max\limits_{y \in Y} c(y)$. From the planner's perspective, taxing such a worker while incentivizing income production involves no trade-offs, as the worker's income choice is highly inelastic. Thus, in the absence of adverse selection, the planner would optimally impose very steep taxes on this highly productive worker. Moreover, Lemma 1 implies that, in the worst-case scenario, a worker with higher baseline productivity θ_i selects an income level leading to higher expected income. Therefore, by introducing some curvature in the tax—i.e., by offering a *non-affine* convex tax schedule—the planner can effectively target more productive workers with higher marginal tax rates, thereby partially replicating the optimal schedule she would implement if workers' baseline productivity were directly taxable.

This intuition is framed primarily in terms of maximizing tax revenue, as most worst-case forces act through the revenue component of the planner's objective: According to Lemma 1, the worker-welfare component is fixed at the value arising under the baseline technology. However, as Theorem 1 demonstrates, any non-progressive tax rule can be replaced with a progressive one that simultaneously improves both *workers' welfare and revenue*. Thus, there is no conflict between the two components of the planner's objective in this argument.

Crucially, our arguments hold *regardless* of the planner's preferences for redistribution among workers or the weight she places on workers' welfare relative to tax revenue. They also do not rely on specific assumptions about the distribution of workers' heterogeneity. In this sense, progressivity is a *qualitative* feature of an optimal tax, independent of the details of the environment. However, as we will argue in Section 3.4, these details do influence the curvature of the optimal tax and, consequently, its *degree of progressivity*.

Welfare-improving tax reform. Building on the above logic, starting from any non-progressive tax, we can design a tax reform—similar to the one outlined in the proof of Theorem 1—that yields a weak improvement in both (worst-case) workers' welfare and tax revenue, while remaining *independent of the specifics of the environment*, such as the choice of the baseline technology $\mathbf{M}^0$. To state the result formally, let $T(y)$ denote the prevailing tax system, and define $c(y) = y - T(y)$. Let $\tilde{c}(y)$ be the *concavification* of $c(y)$—i.e., $\tilde{c}$ is the smallest concave function that majorizes c—and define $\tilde{T}(y) = y - \tilde{c}(y)$.[24] By construction, $\tilde{T}(y)$ is a progressive tax. We refer to the policy of substituting $T(y)$ with $\tilde{T}(y)$ as a *basic progressive tax reform of* T.

Theorem 1* (Basic progressive tax reform). *For any feasible tax T, the basic progressive tax reform of T leads to a weak improvement in both (worst-case) workers' welfare and revenue.*

Theorem 1* provides a detail-free tax policy that improves upon any non-progressive tax system. We refer to it as a *basic* reform because it represents, in a sense, the minimal way for the government to replace the prevailing tax policy with one that yields better worst-case outcomes across all dimensions. However, the planner could achieve even better results by incorporating her knowledge of M_i^0: as shown in Figure 3.1, further (weak) improvements in both workers' welfare and revenue are possible by adopting the alternative consumption rule constructed in the proof of Theorem 1.

Relationship to results in canonical Mirrlees model. We now contrast our findings with the most salient results about tax progressivity in the Mirrleesian framework. As in that framework, suppose that the planner knows the production technology and that this technology corresponds to the one described in Example 1. If the distribution of workers' productivity is assumed to be bounded, it is well known that the optimal tax is flat at the top of the income distribution (Sadka, 1976; Seade, 1977). This result follows from a standard no-distortion-at-the-top argument, which we illustrate in the left panel of Figure 3.2 below. There, $c(y)$ represents a consumption schedule in which the marginal tax rate faced by the most productive worker (in the figure, type $\bar{\theta}$ who produces income $\hat{y}$ in equilibrium) is strictly positive. By moving to a tax schedule such as the one depicted by $\tilde{c}(y)$, where the tax rate above income $\hat{y}$ is strictly lower than that under $c(y)$, the

[24] The result also goes through if we define $\tilde{c}$ as the *monotone concavification* of c—i.e., the smallest *non-decreasing* concave function that majorizes c.

planner can strictly improve total welfare. This is because this change in the tax schedule leads to a weak improvement in type $\bar{\theta}$'s utility and a strict improvement in his income production (from $\hat{y}$ to $\hat{y}'$), without affecting the income choices of less productive workers. Consequently, the tax schedule associated with $c(y)$ cannot be optimal in the Mirrleesian framework with bounded productivity.

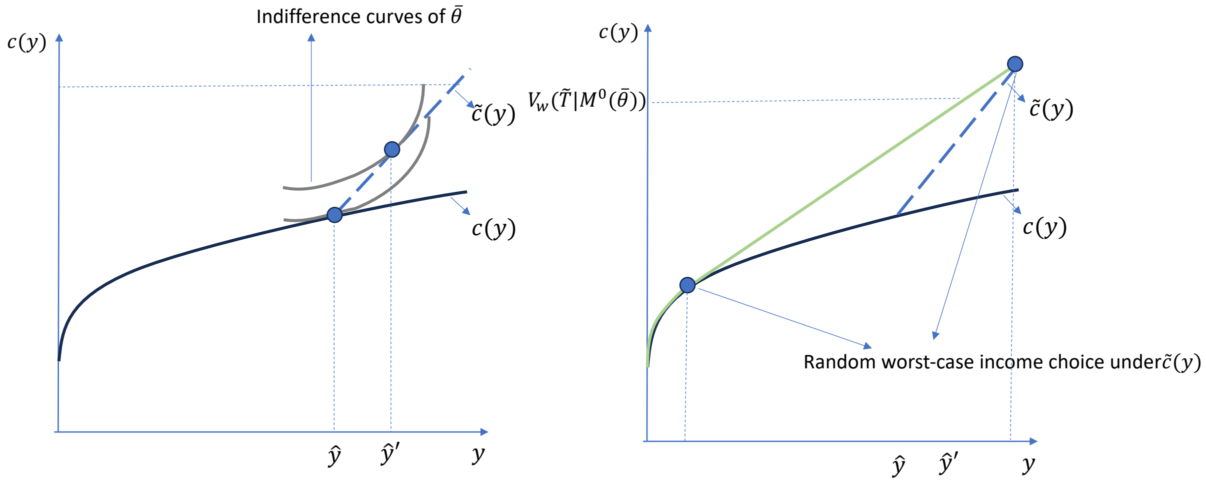

$c(y)$
Indifference curves of $\bar{\theta}$
$\tilde{c}(y)$
$c(y)$
$\hat{y}$
$\hat{y}'$
y
$c(y)$
$V_w(\tilde{T}|M^0(\bar{\theta}))$
$\tilde{c}(y)$
$c(y)$
Random worst-case income choice under$\tilde{c}(y)$
$\hat{y}$
$\hat{y}'$
y

Figure 3.2: A classical no distortion at the top argument in the Mirrlees model with $\overline{\theta} = \max \Theta$ (left panel) and its unraveling in our setting (right panel).

The reason this argument fails in our model is that the perturbation of $c(y)$ we just described does not preserve the concavity of consumption. In our setting, if the non-convex tax schedule associated with $\tilde{c}(y)$ were implemented, then, in the worst case, workers on the higher end of the income distribution would make random income choices, as depicted in the right panel of Figure 3.2, and typically would not lead to an improvement in revenue. This outcome follows from the logic outlined above: strict convexity in the tax schedule implies that, under the worst-case technology, workers find it optimal to take on income risk at a low labor disutility cost, which drives down expected revenue. Consequently, by the argument in Theorem 1, $\tilde{c}(y)$ is dominated by a progressive tax, such as the one illustrated by the green line in the figure. In contrast, in the Mirrleesian model with a known production technology, the worst-case random income choice depicted in the figure is assumed to be unavailable to workers, as income is restricted to being deterministic (or, alternatively, income randomness is assumed to be prohibitively costly for workers).

3.3 Is progressivity necessary?

Given that the worst-case optimal tax need not be unique, we now turn to the question of whether tax progressivity is necessary for optimality. Unsurprisingly, the answer depends on the choice of the relevant set of income realizations over which the convexity of T is required. In Proposition 1, we show that, at the optimum, T must be convex over the range of incomes that may arise in equilibrium, in a sense we define precisely below. Accordingly, the proposition provides an endogenous (T-dependent) range of incomes over which tax progressivity must be satisfied at the optimum. Building on this result, Corollary 1 offers a condition on the model's primitives that ensures this endogenous range coincides with the full income domain Y. Loosely, this condition requires that the set of income realizations feasible under the baseline is sufficiently *rich*.

For any feasible T, let $\mathcal{F}_i^0(T) \subseteq \Delta(Y)$ be the set of optimal income lotteries for worker i under T and M_i^0 subject to favorable tie-breaking, i.e., $\mathcal{F}_i^0(T) \equiv \operatorname*{arg\,max}_{F:(F,\phi)\in X_w(T|M_i^0)} \mathbb{E}_F[T(y)]$, and let $\underline{\mathcal{F}}_i(T) \subseteq \Delta(Y)$ be his income choices under the worst-case scenario as defined by F_i in Lemma 1. Let

$$\mathcal{F}^0(T) \equiv \left\{ \int_{\mathcal{I}} F_i^0 \, di : F_i^0 \in \mathcal{F}_i^0(T), \forall i \in \mathcal{I} \right\}, \quad \underline{\mathcal{F}}(T) \equiv \left\{ \int_{\mathcal{I}} \underline{F}_i \, di : \underline{F}_i \in \underline{\mathcal{F}}_i(T), \forall i \in \mathcal{I} \right\}$$

be the sets of aggregate income distributions that may arise under, respectively, $\mathbf{M}^0$ and the worst-case scenario. Observe that all of the elements in $\mathcal{F}^0(T)$ are outcome-equivalent, in that they lead to the same distribution over worker-welfare and revenue, and the same is true among the elements of $\underline{\mathcal{F}}(T)$. Let $c(y) = y - T(y)$ be the consumption rule associated with $T(y)$, let $\overline{c}(y)$ be the dominating concave consumption rule depicted in Figure 3.1 (which is formally defined by (A.9) in Appendix A.2), and let $\tilde{c}(y)$ be the concavification of $c(y)$.

Proposition 1 (Necessity)**.** *Suppose that $W(\cdot, i)$ is strictly increasing for all $i \in \mathcal{I}$. Then, for any worst-case optimal tax rule T and any $F \in \mathcal{F}^0(T) \cup \underline{\mathcal{F}}(T)$, it holds that $c(y) = \tilde{c}(y) = \overline{c}(y)$ F-almost everywhere.*

As a consequence of Proposition 1 and the definition of $\tilde{c}$, any optimal tax rule must be convex in a neighborhood of every $y \in (0, \overline{y})$ that may arise with positive probability under either the baseline or the worst-case technology. The following Corollary shows that this set of income realizations covers the entire range of income possibilities Y if the baseline technology satisfies the following richness condition: Say that $\mathbf{M}$ satisfies the *richness condition* if, for any selection $\{(F_i, \phi_i)\}_{i\in\mathcal{I}}$ such that $(F_i, \phi_i) \in M_i$ for all $i \in \mathcal{I}$, the distribution $F = \int_{\mathcal{I}} F_i \, di$ has full support on Y.[25]

Corollary 1 (Richness)**.** *Suppose that $W(\cdot, i)$ is strictly increasing for all $i \in \mathcal{I}$. If $\mathbf{M}^0$ satisfies the richness condition, then any worst-case optimal tax rule is convex on Y.*

The result follows immediately from Proposition 1 and the fact that, if $\mathbf{M}^0$ satisfies the richness condition, then F^0 has full support on Y for all $F^0 \in \mathcal{F}^0(T)$ and every feasible T. Intuitively, if $\mathbf{M}^0$ satisfies richness and if $\overline{c}(y) > c(y)$ for a non-degenerate interval of incomes, then $V_w(\overline{c} \mid M_i^0) > V_w(c \mid M_i^0)$ for a positive measure of i. By the arguments given in the proof of Theorem 1, $\overline{c}$ allows the planner to raise weakly higher revenue than c while yielding a strict worker-welfare improvement over c, and therefore c cannot be worst-case optimal.

For example, $\mathbf{M}^0$ satisfies the richness condition if there exists a positive measure of workers for whom every F_i available in M_i^0 has full support on Y. However, non-degenerate lotteries are

[25] In the special case where M_i is constant in i, richness is equivalent to the "full-support" condition in Carroll (2015) which ensures that the optimal contract has to be linear.

not required for the condition to be met. To illustrate this, consider a simple example where $Y = \mathcal{I} = [0,1]$ and $M_i^0 = \{(\delta_i, \phi_i)\}$ for some $\{\phi_i\}_{i\in\mathcal{I}}$. In this case, workers may only make deterministic income choices, but sufficient baseline heterogeneity *across workers* ensures that the aggregate income distribution in equilibrium under $\mathbf{M}^0$ spans the entire set Y.

3.4 Affine taxes and strict progressivity of the optimal policy

Theorem 1 states that there exists a worst-case optimal tax that is *weakly* convex, thereby leaving open the possibility of a worst-case optimal *affine* tax. Within the optimal taxation literature, affine taxes have received significant attention due to their simplicity, especially given that the optimal tax schedule in Mirrlees (1971) can be highly intricate. Furthermore, in the robust contracting model of Carroll (2015) that we generalize, it is known that, in the absence of adverse selection, affine tax schedules are worst-case optimal. With these considerations in mind, this section characterizes the optimal tax schedule within the affine class and identifies conditions under which the optimal tax rule (within the convex class) is *strictly progressive* meaning that it is convex but not affine. Formally, for any convex tax rule T, let $T'(y-)$ and $T'(y+)$ denote the left- and right-directional derivatives, respectively. We define strict progressivity as follows:

Definition 2 (Strict progressivity)**.** $T(y)$ is strictly progressive if it is convex and $T'(0+) < T'(\overline{y}-)$.

Let $T(y) = t + \tau y$ be an affine tax schedule. Let $\mathcal{F}_i^0(\tau) \subseteq \Delta(Y)$ be the set of worker i's chosen income lotteries under M_i^0,[26] and let $F_i^0(\tau) \in \mathcal{F}_i^0(\tau)$ and $y_i^0(\tau) = \mathbb{E}_{F_i^0(\tau)}[y]$. Let $\phi_i^0(\tau)$ be such that $(F_i^0(\tau), \phi_i^0(\tau))$ is optimal under M_i^0 and τ. Let $w_i(t,\tau) = W_1(-t + (1-\tau)y_i^0(\tau) - \phi_i^0(\tau), i)$, which is the social welfare weight of worker i under T. With some abuse of notation, we write $V_P(t,\tau)$ to refer to worst-case welfare under the affine schedule T. We say that the tax schedule $T(y)$ is *worst-case optimal within the affine class* if (t,τ) solves[27]

$$\max_{t\in\mathbb{R}_-,\tau\in[0,1]} V_P(t,\tau).$$

Proposition 2 provides a condition (equation (3.4)) under which any such rule is strictly suboptimal, and therefore any worst-case optimal progressive tax has to be *strictly* progressive. For simplicity, we drop the dependence of the above objects with respect to (t,τ), which are understood to be evaluated at their optimal values (t^*,τ^*).

Proposition 2 (Suboptimality of affine taxes)**.** *There exists a tax schedule $T_a(y) = t^* + \tau^* y$ that is worst-case optimal within the affine class. For any such T_a with $\tau^* < 1$, if for some $\tilde{y} \in Y$, it holds that*

$$\int_{\mathcal{I}} \left[\left(w_i + \alpha\frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F_i^0}[\min\{\tilde{y}-y,0\}] - \alpha\mathbf{I}(y_i^0 > \tilde{y})\frac{\tilde{y}-y_i^0+\phi_i^0/(1-\tau^*)}{1-\tau^*}\right] di > 0, \qquad (3.4)$$

[26] This choice is independent of t.

[27] By building on the proof of Theorem 1, it is straightforward to show that it is without loss of optimality to focus on double-monotone affine tax rules, which requires that $\tau \in [0,1]$.

then there exists a strictly progressive tax rule that yields strictly higher worst-case welfare than T_a.

To better understand the logic behind condition 3.4, it will be convenient to first characterize the optimal values of (t^*, τ^*) and discuss their derivation. To simplify exposition, we focus on the case in which (t^*, τ^*) can be obtained through first-order conditions, for which we introduce the following *free-option* assumption.

Assumption 3 (Free option)**.** For each $i \in \mathcal{I}$, there exists $F_i \in \Delta(Y)$ such that $(F_i, 0) \in M_i^0$.

Our next result characterizes a worst-case optimal tax within the affine class.

Proposition 3 (Worst-case optimal affine tax)**.** *Suppose that Assumption 3 holds. Any tax function $T_a(y) = t^* + \tau^* y$ that is worst-case optimal within the affine class satisfies the condition*

$$\int_{\mathcal{I}} (w_i(t^*, \tau^*) - \alpha)\, di \leq 0, \quad \textit{with equality if } t^* < 0. \tag{3.5}$$

Furthermore, if there exists a pair $(\tau^o, t^o) \in [0,1) \times \mathbb{R}_-$ that satisfies (3.5) and the following conditions:

$$\tau^o = \max\left\{1 - \sqrt{\frac{\alpha \int_{\mathcal{I}} \phi_i^0(\tau^o)\, di}{\int_{\mathcal{I}} (\alpha - w_i(t^o, \tau^o)) y_i^0(\tau^o)\, di}}, 0\right\}, \quad \textit{and} \tag{3.6}$$

$$V_P(t^o, \tau^o) \geq V_P(t, 1) \quad \textit{for all } t \leq 0, \tag{3.7}$$

then (t^o, τ^o) is worst-case optimal within the affine class. Otherwise, the pair $(t(1), 1)$, where $t(1)$ satisfies (3.5) for $\tau^* = 1$, is worst-case optimal within this class.

Condition (3.5), which pins down the lump-sum payment $t^* = T(0)$ for every value of τ^*,[28] is intuitive: If $t^* < 0$ (workers receive a positive lump-sum payment), then the transfer amount has to be such that the marginal benefit of redistributing an extra dollar—which is given by the average social welfare weight—is equal to the marginal cost—which is represented by the marginal value of public funds, α. Otherwise, the non-negativity constraint on consumption is binding and optimal lump-sum payments are equal to zero. An analogous condition arises in the canonical model if we substitute α for the Lagrange multiplier associated with the government's budget constraint.

Next, consider the worst-case optimal marginal tax rate τ^*. Let us focus on the most interesting case in which $\tau^* \in (0,1)$. Applying Lemma 1, the worst-case income choice of worker i under the marginal tax rate τ, whose mean we denote by $y_i(\tau) \in Y$, is such that

$$V_w(T|M_i^0) = (1-\tau) y_i^0(\tau) - t - \phi_i^0(\tau) = (1-\tau) y_i(\tau) - t \iff y_i(\tau) = y_i^0(\tau) - \frac{\phi_i^0(\tau)}{1-\tau}, \tag{3.8}$$

and worst-case consumption is equal to $V_w(T|M_i^0)$.

Following the discussion in Saez (2001), we can use (3.8) to informally decompose the first-order worst-case welfare effect of changing the tax rate τ into two parts. First, the *mechanical*

[28] Assumption 2 ensures that a solution exists for every $\tau^* \in [0,1]$.

effect reflects that, holding workers' income choices fixed, their worst-case consumption and tax payments change proportionally to $y_i^0(\tau)$. This yields the mechanical effect on total welfare:

$$\int_{\mathcal{I}} (\alpha - w_i(t,\tau)) y_i^0(\tau)\, di.$$

Second, there is a *worst-case behavioral* response, which captures the fact that workers' worst-case income choice is affected by the change in τ. According to (3.8), the first-order effect on worker i's income production is equal to $-\phi_i^0(\tau)/(1-\tau)^2$.[29] Thus, the behavioral welfare change is equal to

$$-\alpha \int_{\mathcal{I}} \frac{\phi_i^0(\tau)}{(1-\tau)^2}\, di.$$

Setting the sum of the two effects equal to zero gives the tax rate in (3.6). The reason why we need to verify the additional condition (3.7) is that, due to workers' favorable tie-breaking, $V_P(\tau)$ has a discontinuity at $\tau = 1$.

We note that the expression for the worst-case behavioral response is qualitatively different—i.e., it depends on different model primitives—compared to its analog in the canonical framework studied by Saez (2001). Unlike that setting, our behavioral effect accounts for the fact that the *worst-case production technology is not fixed as the tax rate changes.* The key implication is that the behavioral effect is no longer described in terms of labor supply elasticity, as in Saez (2001). Instead, in our model, the behavioral response is governed by the *level* of workers' labor disutility under the baseline technology. Intuitively, $\phi_i^0(\tau)$ captures the extent to which revenue can be lowered through an adversarially chosen technology that introduces a new "low-cost" income option into M_i^0.

By applying a similar logic, we can derive condition (3.4) in Proposition 2, which rules out the possibility that an affine tax is (fully) worst-case optimal. For the sake of exposition, suppose momentarily that workers' optimal income choice under T_a and M_i^0—i.e., $\mathcal{F}_i^0(\tau^*)$—is single-valued and deterministic for all $i \in \mathcal{I}$. Consider the perturbation around $T_a(y)$ given by $T^\varepsilon(y) \equiv T_a(y) + \varepsilon \max\{y - \tilde{y}, 0\}$, with $\tilde{y} \in (\inf_{i\in\mathcal{I}} y_i^0(\tau^*), \sup_{i\in\mathcal{I}} y_i^0(\tau^*))$ and $\varepsilon > 0$. In words, $T^\varepsilon(y)$ is a strictly progressive perturbation that involves increasing the tax rate by ε above the income level $\tilde{y}$. We show in Appendix B (Lemma 6) that the planner's objective is directionally differentiable with respect to ε and that the marginal effect on total welfare as $\varepsilon \downarrow 0$ is given by the left-hand-side of (3.4). Under the simplifying assumptions made here, that condition reduces to

$$\int_{\mathcal{I}} \mathbb{I}(y_i^0(\tau^*) > \tilde{y}) \left((\alpha - w_i(t^*,\tau^*))(y_i^0(\tau^*) - \tilde{y}) - \alpha \frac{\phi_i^0(\tau^*)}{(1-\tau^*)^2} \right) di > 0. \tag{3.9}$$

Intuitively, an increase in the tax rate above $\tilde{y}$ only affects workers whose baseline income $y_i^0(\tau^*)$ is above this cutoff. Within that region of the income distribution, an analogous intuition to the

[29] We show in the proof of Proposition 2 that the behavioral effect through a change in $(y_i^0(\tau), \phi_i^0(\tau))$ is of second order.

one given in Proposition 3 yields that the mechanical welfare effect is

$$\int_{\mathcal{I}} \mathbb{I}(y_i^0(\tau^*) > \tilde{y})(\alpha - w_i(t^*, \tau^*))(y_i^0(\tau^*) - \tilde{y})\, di,$$

and the worst-case behavioral one is

$$-\alpha \int_{\mathcal{I}} \mathbb{I}(y_i^0(\tau^*) > \tilde{y})\phi_i^0(\tau^*)/(1 - \tau^*)^2\, di.$$

Thus, if condition (3.9) is satisfied, $T^{\varepsilon}(y)$ yields a strict improvement over $T_a(y)$.

Recall from our discussion of the proof of Theorem 1 that an affine tax is optimal in the benchmark with a single type. Thus, a minimal amount of heterogeneity with respect to i is required for (3.4) to hold. Indeed, the definition of τ^* ensures that (3.4) is never satisfied whenever M_i^0 is constant in i.

A natural question is whether the negation of (3.4) is sufficient to ensure that an affine tax rule is optimal. Unfortunately, the perturbation argument that we provide in the proof does not enable us to derive sufficient conditions for the optimality of $T_a(y)$. The reason is that the planner's objective is neither differentiable with respect to all perturbations nor concave in T. In the binary-type model that we study in Section 4.1, condition (3.4) is indeed necessary and sufficient to rule out optimality of an affine tax.

Example 1 (cont.). To further illustrate the intuition behind (3.9), consider the special case where M_i^0 is defined as in Example 1, so that $y_i^0(\tau^*)$ is increasing in θ_i. In this case, condition (3.9) is always satisfied if $\tau^* < 1$ and there exists a cutoff $\tilde{\theta}$ such that $\mathbb{E}[\phi_i^0(\tau^*)|\theta_i > \tilde{\theta}]$ is arbitrarily close to zero and $\mathbb{E}[\alpha - w_i(t^*, \tau^*)|\theta_i > \tilde{\theta}] > 0$—i.e., if the cost of income production is negligible for the most skilled workers, and the government assigns relatively low social welfare weight to them. The first condition holds when the worst-case income elasticity of the most skilled workers is relatively low. In this scenario, the planner can impose heavy taxes on these workers while still incentivizing high income production. This logic aligns with our intuition in Section 3.2 that an increasing slope of the revenue-maximizing worker-specific affine tax with respect to baseline income $y_i^0(\tau)$ pushes for *strict* convexity of the optimal tax.

4 Optimal tax under one-dimensional finite-support types

To make progress in characterizing optimal tax schedules, we specialize the model in three main ways in this section. First, as in the intuitive proof of Theorem 1 in Section 3, we assume that workers have a one-dimensional type governing their baseline set of actions. Second, we assume that the set of possible type realizations is finite. Third, we impose assumptions on workers' baseline labor disutility that ensure that, for any progressive tax schedule, workers' baseline income choices are deterministic and increasing in their type.

The goal of this section is twofold. First, we illustrate how optimal taxes can be computed in our framework once all the information that the planner may have about the production technology is incorporated into the problem—this information is captured by the details of the baseline technology $\mathbf{M}^0$. Second, the closed-form solutions derived in this section can be applied to analyze how the details of the economy affect the curvature and other features of the tax schedule.

Framework. We assume that each worker i has a type θ_i that takes on values in $\Theta \equiv \{\theta^1, ..., \theta^n\} \subseteq \mathbb{R}$.[30] We label the elements of Θ so that $\theta^1 < ... < \theta^n$. We assume that, for each $\Theta' \subseteq \Theta$, the set $\{i \in \mathcal{I} : \theta_i \in \Theta'\}$ is Lebesgue-measurable. For $j \in N$, we write $p^j = \int_{\mathcal{I}} \mathbb{I}(\theta_i = \theta^j)\, di$ which is the mass of workers with type θ^j, and without loss assume that $p^j \in (0,1)$ for all j. We assume that the distribution $(p^1, ..., p^n)$ is known by the social planner. Furthermore, we assume that there exists a correspondence $M^0 : \Theta \rightrightarrows \mathcal{M}$ such that the baseline technology is described by $M^0(\theta_i) = M_i^0$ for all $i \in \mathcal{I}$. We impose the following assumptions on M^0.

Assumption 4 (Baseline risk preference)**.** For every $\theta \in \Theta$ and $(F, \phi) \in M^0(\theta)$, there exists $\phi' \leq \phi$ such that $(\delta_{\mathbb{E}_F[y]}, \phi') \in M^0(\theta)$, with the inequality being strict whenever $F \neq \delta_{\mathbb{E}_F[y]}$.

Assumption 4 says that, for every non-degenerate income choice that is available under the baseline, there is also available a degenerate lottery with the same mean that is strictly less costly. This assumption implies that, for any weakly concave consumption schedule, workers' income choice under the baseline technology is always deterministic. Given this, we define the effective cost of producing deterministic income $y \in Y$ as $\Phi(y, \theta) \equiv \min\{\phi : (\delta_y, \phi) \in M^0(\theta)\}$, with the convention that $\Phi(y, \theta) = +\infty$ whenever we are minimizing over the empty set.

Assumption 5 (Baseline labor disutility)**.** $\Phi(y, \theta)$ satisfies the following conditions:

(i) *Groundedness and positivity:* For every $\theta \in \Theta$, $\Phi(0, \theta) = 0$ and $\Phi(y, \theta) > 0$ for all $y > 0$;

(ii) *Non-triviality:* There exists $\theta \in \Theta$ and $y \in Y$ such that $y - \Phi(y, \theta) > 0$;

(iii) *Decreasing differences:* For every $y, y' \in Y$ and $\theta, \theta' \in \Theta$ such that $y' > y$ and $\theta' < \theta$, it holds that

$$\Phi(y', \theta') + \Phi(y, \theta) \geq \Phi(y', \theta) + \Phi(y, \theta'),$$

with strict inequality if $\Phi(y', \theta) + \Phi(y, \theta') < +\infty$.

The first part of Assumption 5 is a strengthening of the free-option assumption (Assumption 3). The second part ensures that there exists a positive measure of workers who can produce positive welfare under the baseline. The third part is a standard decreasing differences condition, stating that an increase in income production is more costly for lower types.[31]

Although stylized, this version of the model allows us to derive closed-form solutions for the optimal tax schedule and to perform comparative statics on the model's primitives. A possible interpretation of this framework is one where the social planner has some coarse information about

[30] A similar analysis would also apply if the type space were continuous but the planner was restricted to using a piecewise affine tax rule with finitely many income brackets—as is always observed in practice.

[31] We write the condition as a sum, instead of using the more common condition $\Phi(y', \theta') - \Phi(y, \theta') > \Phi(y', \theta) - \Phi(y, \theta)$, in order to avoid dealing with the operation $+\infty - (+\infty)$.

workers (e.g., their wage or level of education), and can use that information to infer workers' baseline labor opportunities through $M^0(\theta_i)$.

Our assumptions on $\Phi(y,\theta)$ are weaker than those typically imposed in the standard Mirrleesian framework—e.g., we do not impose smoothness, monotonicity or convexity of $\Phi(\cdot,\theta)$, or require workers' available income choices to be an interval. Yet, Assumptions 4 and 5 yield a simple characterization of workers' equilibrium choice under the baseline technology, which is similar to the equilbrium structure that arises under the standard model. Namely, workers' equilibrium income choices are deterministic, and income and workers' indirect utilities are increasing in θ.

Lemma 2 (Baseline equilibrium)**.** *Suppose that Assumptions 4 and 5 hold. For any convex tax rule, $T(y)$, the income choice for type θ^j under the baseline technology, (F_0^j,ϕ_0^j), satisfies the following properties:*

(i) $F_0^j = \delta_{y_0^j}$, where $y_0^j \equiv \mathbb{E}_{F_0^j}[y]$;

(ii) For all $j > j'$, $y_0^j \geq y_0^{j'}$ and $V_w(T|M^0(\theta^j)) \geq V_w(T|M^0(\theta^{j'}))$.

Simple tax schedules. As a final preliminary toward characterizing the worst-case optimal tax, we introduce a definition that allows us to focus on taxes that do not contain any "artificial" income brackets, meaning income brackets that are never chosen by workers in equilibrium and that play no role in shaping workers' incentives.

As a first step toward this, we specialize to *at-most n-piece affine tax schedules*. The fact that this restriction entails no loss of optimality can be deduced from the intuitive proof of Theorem 1 discussed in Section 3. There, any consumption schedule that uses more than n different marginal rates is weakly dominated by the at-most n-piece affine consumption schedule obtained by taking the pointwise minimum across the n relevant type-specific affine functions described in the proof.

Thus, for any $k \leq n$, $\hat{\mathbf{y}} \in Y^{k-1}$, $t \in \mathbb{R}_-$ and $\boldsymbol{\tau} \in [0,1]^k$, we define

$$T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau}) = \begin{cases} t + \tau^1 y, & \text{if } y \in [0,\hat{y}^1] \\ t + \tau^1\hat{y}^1 + \tau^2(y-\hat{y}^2), & \text{if } y \in [\hat{y}^1,\hat{y}^2] \\ ... \\ t + \sum_{l=1}^{k-1}\tau^l(\hat{y}^l - \hat{y}^{l-1}) + \tau^k(y-\hat{y}^l), & \text{if } y \in [\hat{y}^{k-1},\bar{y}], \end{cases} \tag{4.1}$$

to be the tax schedule with k income brackets defined by the cutoffs $(\hat{y}^1,...,\hat{y}^{k-1})$, a lump-sum transfer equal to t, and bracket-specific marginal tax rates given by $(\tau^1,...,\tau^k)$. We write $c^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau}) = y - T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau})$. For convenience, we will denote $\hat{y}^0 = 0$ and $\hat{y}^k = \bar{y}$. Let $(y_0^1,...,y_0^n) \in Y^n$ and $(y^1,...,y^n) \in Y^n$ denote the vectors of type-dependent expected income production under, respectively, the baseline and the worst-case technology from Lemma 1. Set $y_0^{n+1} = y^{n+1} = \bar{y}$. Throughout the rest of this section, we will use boldface font to refer to any vector of the form $\mathbf{x} = (x^1,..,x^k) \in \mathbb{R}^k$.

Our second simplification of the space of relevant taxes is twofold: First, our definition of an income partition $\hat{\mathbf{y}} \in Y^{k-1}$ does not rule out the possibility that two different partition cells have the same marginal tax rate. We get rid of redundant partition elements by merging all income brackets with the same rate into one. Second, we know from the proof of Theorem 1 that it is without loss of optimality to restrict attention to tax schedules with a marginal tax rate that is constant on each interval $[y^j, y^{j+1}]$ for each $j \in \{0, ..., n\}$.

We refer to a tax schedule satisfying these conditions as *simple*. A formal definition of simple tax schedules is provided in Online Appendix OA.1. In words, for any feasible tax T, let $\overline{T}$ be the welfare-dominating tax schedule that obtains from applying the process described in Figure 3.1. T is simple if $T = \overline{T}$. Our next result states that, without loss of optimality or existence, we may restrict attention to the space of simple taxes.

Proposition 4 (Optimality of simple taxes)**.** *There exists a worst-case optimal tax that is simple.*

Building on this finding, Proposition 5 provides necessary conditions for worst-case optimality of a simple tax schedule. These conditions allow to compute the optimal values of $(t, \boldsymbol{\tau})$. Proposition 4 implies that, paired with an optimally chosen income partition $\hat{\mathbf{y}}$, the conditions in Proposition 5 are also sufficient for worst-case optimality. To simplify exposition, we introduce the assumption that the social planner's welfare function $W(u, i)$ is linear in u—so that preferences for redistribution are captured by an exogenous Pareto weight—and that she is inequality averse.

Assumption 6 (Preference for redistribution)**.** There exists a function $i \to \omega(i)$ such that $W(u, i) = \omega(i)u$ for all $i \in \mathcal{I}$. Moreover, $\omega^j \equiv \frac{1}{p^j} \times \int_{\mathcal{I}} \mathbb{I}(\theta_i = \theta^j)\omega(i)\, di$ is strictly decreasing in j.

Proposition 5 (Optimal rates)**.** *In any worst-case optimal tax schedule $T^k(y; \hat{\mathbf{y}}, t, \boldsymbol{\tau})$ that is simple, the marginal tax rates $\boldsymbol{\tau}$ are defined recursively by*

$$\tau^k = \begin{cases} 0, \quad \textit{if} \quad \mathbb{E}[\alpha - \omega^j | y_0^j \geq \hat{y}^{k-1}] = 0, \\ \max\left\{0, 1 - \sqrt{\frac{\alpha \mathbb{E}[\Phi(y_0^j, \theta^j) | y^j \geq \hat{y}^{k-1}] \Pr(y^j \geq \hat{y}^{k-1})}{\mathbb{E}[(\alpha - \omega^j)(y_0^j - \hat{y}_{k-1}) | y_0^j \geq \hat{y}^{k-1}] \Pr(y_0^j \geq \hat{y}^{k-1})}}\right\}, \quad \textit{otherwise;} \end{cases} \tag{4.2}$$

And, if $k > 1$, then for $l = 1, ..., k-1$:

$$\tau^l = \begin{cases} 0, \quad \textit{if} \quad \mathbb{E}[\alpha - \omega^j | y_0^j \geq \hat{y}^{l-1}] = 0, \\ \max\left\{0, 1 - \sqrt{\frac{\alpha \mathbb{E}\left[\Phi(y_0^j, \theta^j) - \sum_{h=l}^{k-1} \mathbb{I}(y_0^j \geq \hat{y}^h)(1 - \tau^{h+1})(\min\{\hat{y}^{h+1}, y_0^j\} - \hat{y}^h) | y^j \in [\hat{y}^{l-1}, \hat{y}^l]\right] \Pr(y^j \in [\hat{y}^{l-1}, \hat{y}^l])}{\mathbb{E}[(\alpha - \omega^j)(\min\{y_0^j, \hat{y}^l\} - \hat{y}^{l-1}) | y_0^j \geq \hat{y}^{l-1}] \Pr(y_0^j \geq \hat{y}^{l-1})}}\right\}, \\ \quad \textit{otherwise.} \end{cases} \tag{4.3}$$

As an application of Proposition 5, Corollary 2, states that the optimal tax rate at the top converges to full taxation if the highest type's marginal cost of producing income converges to zero. Consider a sequence of labor disutility functions $\{\Phi_N(y, \theta)\}_{N=1}^{+\infty}$ such that $\Phi_N(y, \theta^n) \to 0$ for all

$y \in Y$. Under Assumption 5 (groundedness), this is equivalent to the *marginal* labor disutility of θ^n converging to 0. For simplicity, we keep $\Phi_N(y, \theta^j)$ constant in N for all $j \in \{1, ..., n-1\}$.[32] Suppose further that, for all $j \in \{1, ..., n-1\}$, there exists $\tilde{y} < \bar{y}$ such that $y - \Phi_N(y, \theta^j) < 0$ for all $y \geq \tilde{y}$, which implies that the baseline income choice of all workers with type $\theta < \theta^n$ is bounded away from $\bar{y}$. Let $T_N(y)$ be a simple worst-case optimal tax schedule under $\Phi_N(y, \theta)$. By compactness of the space of simple tax schedules, we can without loss assume that $T_N(y)$ has a (simple) limit, which we denote by $T_\infty(y)$ and its right-derivative by $T'_\infty(y)$.

Corollary 2 (Full taxation at the top)**.** *There exists $y^* < \bar{y}$ such that $T'_\infty(y) = 1$ for all $y \geq y^*$.*

By definition of a simple tax schedule, type θ^n's worst-case income is above y^* with strictly positive probability and by Corollary 2, those high income realizations are fully marginally taxed. This result stands in contrast with zero marginal taxation at the top in the canonical framework. It follows from the assumption that the planner is inequality averse (Assumption 6) together with our recurring intuition that, if a worker is extremely productive, the planner is able to tax him heavily and in that way raise higher revenue with little allocative distortion.

4.1 Application: Binary types

To illustrate Proposition 5 and highlight additional features of the worst-case optimal tax, while still accounting for the frictions arising from worker heterogeneity, let us consider the case with binary types—i.e., $n = 2$. The formal results and extra computations that are specific to this binary setting can be found in Online Appendix OA.3.

First, we discuss conditions for worst-case optimality of an affine tax. Let $T_a(y) = t^* + \tau^* y$ be any worst-case optimal tax within the affine class from Proposition 3, and let $y_0^j(\tau)$ be the mean of a baseline income choice of a worker with type θ^j under an affine tax with marginal tax rate equal to τ. As shown in Proposition OA.1, in the binary setting, the affine tax $T_a(y)$ is (fully) worst-case optimal if and only if

$$(\omega^2 - \alpha)(y_0^2(\tau^*) - y_0^1(\tau^*)) + \frac{\alpha \Phi(y_0^2(\tau^*), \theta^2)}{(1-\tau^*)^2} \geq 0. \tag{4.4}$$

Condition (4.4) is a special case of the condition in Proposition 2 with $\tilde{y} = y_0^1(\tau^*)$. Essentially, this condition ensures that it is not beneficial to perturb the optimal affine tax by slightly increasing the tax rate above the baseline income level of the low type. With binary types, this is the only *relevant* perturbation, and thus condition (4.4) is also sufficient for worst-case optimality of T_a.

Figure 4.1 depicts the worst-case optimal tax schedule for a case in which (4.4) is violated and therefore the optimal tax is strictly progressive. At the optimum, the kink is placed at the baseline income choice of the lower type—i.e., $\hat{y} = y_0^1$. The intuition for this result can be understood as follows. By Lemma 2, increasing the kink above y_0^1 only leads to a flat reduction in the tax

[32] Everything goes through if instead we assume that there is a cutoff $\bar{j} \in \{2, ..., n\}$ such that $\Phi_N(y, \theta^j) \to 0$ for all $j \in \{\bar{j}, ..., n\}$.

paid by type θ^2, which by Assumption 6, the planner values according to $p^2(\omega^2 - \alpha) < 0$. As a consequence, the income choice of the lower type is distorted in equilibrium: Even if there is zero marginal taxation below y_0^1, type θ^1 would like to produce more income if the marginal tax rate was everywhere constant and equal to τ^1. We also show that the distortion in y_0^1 is proportional to the probability of the high type, and therefore disappears in the absence of worker heterogeneity.

The optimal marginal tax rates (τ^1, τ^2) are given by

$$\tau^1 = \begin{cases} 0, & \text{if } \mathbb{E}[\omega^j] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{p^1 \alpha \Phi(y_0^1, \theta^1)}{(\alpha - \mathbb{E}[\omega^j]) y_0^1}}\right\}, & \text{if } \mathbb{E}[\omega^j] < \alpha, \end{cases} \qquad \tau^2 = 1 - \sqrt{\frac{\alpha \Phi(y_0^2, \theta^2)}{(\alpha - \omega^2)(y_0^2 - y_0^1)}}.$$

As in Section 3.4, we can interpret the expressions for τ^1 and τ^2 in light of the mechanical and worst-case behavioral effects of perturbing the marginal taxes. A perturbation of the tax rate above y_0^1, which only affects type θ^2, leads to a total welfare effect that can be decomposed as

$$\underbrace{p^2(\alpha - \omega^2)(y_0^2 - y_0^1)}_{\text{mechanical}} - \underbrace{p^2 \alpha \frac{\Phi(y_0^2, \theta^2)}{(1 - \tau^2)^2}}_{\text{behavioral}}.$$

Setting the total welfare effect equal to zero gives the expression for τ^2. By a similar logic, a small increase in τ^1 leads to a mechanical increase of the tax payment made by *all workers*, with an effect on total welfare which is equal to $(\alpha - \mathbb{E}[\omega^j]) y_0^1$. On the other hand, the change only affects the income choice of the low type, and therefore the behavioral welfare loss of increasing τ^1 is given by $p^1 \alpha \frac{\Phi(y_0^1, \theta^1)}{(1-\tau^1)^2}$. This explains the formula for τ^1.

Another pattern illustrated in Figure 4.1, which holds more generally, is that the optimal tax schedule *does not* coincide with the pointwise minimum of the type-specific optimal affine tax in a benchmark with observable types. This contrasts with the intuition suggested by the proof of Theorem 1 and highlights the role of the Bayesian component in the planner's objective: Even though the 'inf' part of the planner's problem can be computed type-by-type, independent of the distribution of types (Lemma 1), the tax rule attaining the 'sup' involves a trade-off, as it must tax different types using the same tax schedule. Consequently, as the figure shows, the optimal schedule is shaped by the need to provide the high type with an information rent.

Comparative statics. We now use this simple setting to study the comparative statics of the optimal tax with respect to the model's primitives. We provide numerical comparative statics results in this section, and we show in Corollaries OA.1 and OA.2 in Online Appendix OA.3 that these results hold analytically under additional assumptions that ensure that the maximizer is unique and differentiable with respect to the relevant parameters.

As expected, the optimal tax burden is higher when the planner's value for revenue, α, increases. This result is graphically depicted for the binary-type case in Figure 4.2. The figure also illustrates that tax progressivity, as measured by $\tau^2 - \tau^1$, is not necessarily monotone in α.

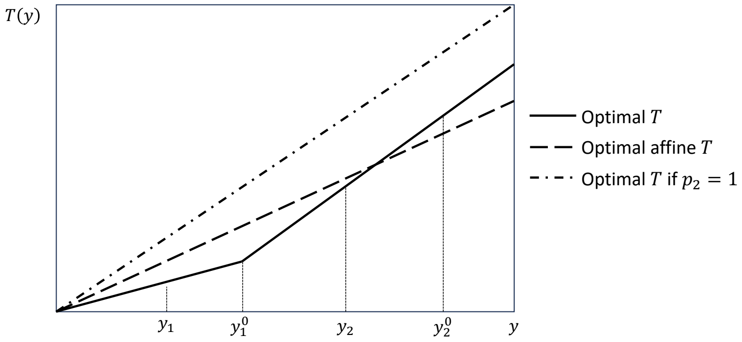


Figure 4.1: Worst-case optimal tax schedule under binary types: An example where strict progressivity is optimal.

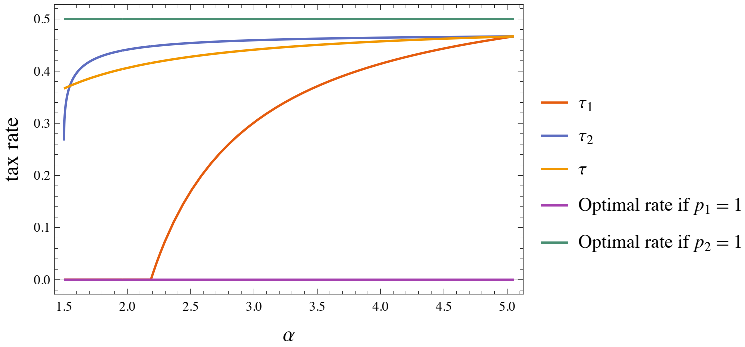


Figure 4.2: Tax rates as a function of α, in a region where strict progressivity is optimal ((4.4) does not hold). Parameter values: $\Phi(y, \theta) = y^2/(2\theta^2)$, $\theta^1 = 1$, $\theta^2 = 4$, $p^1 = 0.75$, $\omega^1 = 2$, $\omega^2 = 0$.

A similar conclusion is obtained when looking at how tax rates evolve with the planner's preference for redistribution, as measured by the Pareto weights ω^j (Figure 4.3). An increase in ω^2—i.e., a decrease in the planner's inequality aversion—leads to a decrease in both τ^1 and τ^2, and therefore an overall downward shift in the optimal tax schedule. Intuitively, the mechanical effect of slightly increasing τ^1 and τ^2 is, respectively $(\alpha - \mathbb{E}[\omega^j])y_0^1$ and $p^2(\alpha - \omega^2)(y_0^2 - y_0^1)$, both of which are decreasing in ω^2.[33] Consequently, a higher ω^2 reduces the relative gain from raising revenue and therefore favors the use of lower marginal taxes over the entire income distribution.

Finally, as depicted in Figure 4.3, the overall effect of changing ω^2 on $\tau^2 - \tau^1$ is ambiguous. Thus, without additional assumptions about the primitives, no definitive relationship emerges between the planner's preference for redistribution and the steepness of marginal tax rates.[34] In Section 5, we further discuss the potential challenges in establishing a general comparative statics result that links tax progressivity to the planner's preference for redistribution.

[33] We show in Online Appendix OA.3 that the behavioral effect is of second order.

[34] However, we show in Corollary OA.2 that if the planner's *average* valuation of workers' welfare, $\mathbb{E}[\omega^j]$, is sufficiently high, the intuitive comparative statics hold: the progressivity of the tax $(\tau^2 - \tau^1)$ increases with the degree of inequality aversion $(\omega^1 - \omega^2)$. Specifically, if we assume, as in the canonical Mirrleesian model, that $\mathbb{E}[\omega^j] = \alpha$, then $\tau^2 - \tau^1$ indeed (locally) increases with the planner's preference for redistribution.

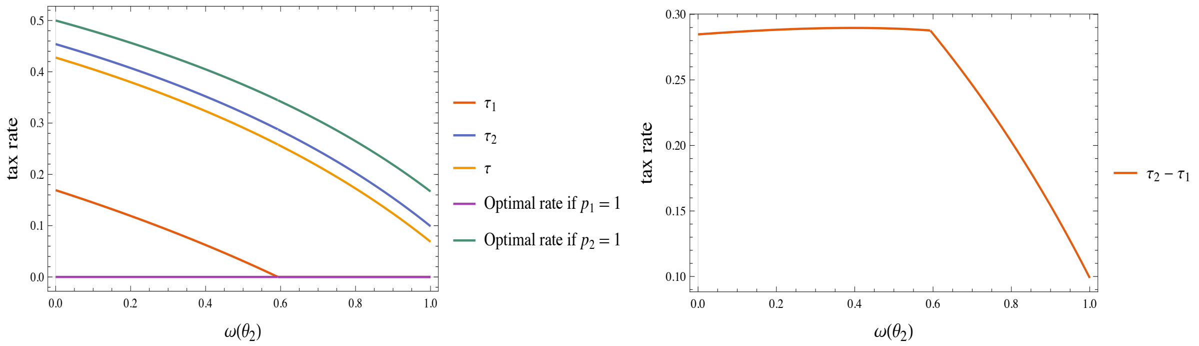


Figure 4.3: Tax rates as a function of ω^2, in a region where strong progressivity is optimal ((4.4) does not hold). Parameter values: $\Phi(y, \theta) = y^2/(2\theta^2)$, $\theta^1 = 1$, $\theta^2 = 4$, $p^1 = 0.75$, $\omega^1 = 2$, $\alpha = 2.5$.

5 Discussion and Extensions

We studied the design of optimal income taxes in the presence of limited information about the technology and robust concerns by the government. Using this approach, we established that the optimal tax rule is progressive, and that this result holds regardless of the government's attitudes toward inequality or the specifics of the distribution of workers' types. More generally, our model can be viewed as an instance of robust principal-agent contracting under moral hazard and adverse selection, in which case our main result provides a foundation for the use of concave contracts.

To maintain tractability, our model relies on a set of stylized assumptions about workers' utility functions, the planner's information structure, and the set of available mechanisms. These assumptions enable us to demonstrate a key qualitative feature of the optimal tax. To conclude, we discuss their limitations and the extent to which our results hold when these assumptions are relaxed.

Other structures of uncertainty about the production technology. The uncertainty faced by our social planner regarding the production technology $\mathbf{M}$ is very rich. Our main result continues to hold even if the planner considers a smaller class of technologies, specifically those of the form $M_i = M_i^0 \cup (F_i, \phi_i)$ with $(F_i, \phi_i) \in \Delta(Y) \times \mathbb{R}_+$. In other words, this assumes that the planner entertains the possibility that workers have a single additional income opportunity added to their baseline choice set. As shown in Lemma 1, in the worst case, this additional income choice minimizes revenue and does not improve the worker's welfare relative to the baseline. Moreover, the structure of the problem remains unchanged if, instead of minimizing over the entire set $\Delta(Y)$, the worst case is computed across all distributions on Y with *binary* support. Consequently, our findings do not require workers' income to be drawn from highly complex lotteries.

In an earlier version of this paper, we also showed that our results hold in the simpler setting where workers have a one-dimensional type $\theta \in \Theta$, any technology is described by a correspondence $M : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$, and the planner considers a *smaller* set of plausible correspondences, specifically those that are continuous and monotone. That framework also accommodates uncertainty about the distribution of θ, in addition to uncertainty about M, without affecting the main results.

Overall, the set of plausible technologies assumed in this paper is a natural starting point, which has received considerable attention in the robust contracting literature. There are certainly other information structures the planner may encounter in practice that could yield different insights. For instance, one could consider a case where the planner faces *a small amount of uncertainty*, contemplating only correspondences within a neighborhood of the baseline $\mathbf{M}^0$. If, moreover, $\mathbf{M}^0$ were assumed to satisfy the typical assumptions of a Mirrlees economy, we would expect the resulting outcomes to be similar to those in the canonical framework.

Screening. We assumed that the planner is restricted to offering a single tax rule for the entire population of workers. This simplifying assumption is, in many cases, realistic, but it can result in a loss of optimality. Here, we examine what happens if the planner can offer workers a *menu of tax rules*, allowing each worker, knowing M_i, to self-select into their most preferred option.

Formally, suppose that the planner can offer any compact menu of feasible tax rules $\mathcal{T} \subseteq \mathbf{T}_f$. When facing $\mathcal{T}$ and the choice set M_i, worker i chooses a tax and an income choice to solve

$$V_w(\mathcal{T}|M_i) = \max_{T \in \mathcal{T}, (F,\phi) \in M_i} \{\mathbb{E}_F[y - T(y)] - \phi\}.$$

Given workers' income and tax choices, we let $V_P(\mathcal{T}|M)$ denote total welfare under the menu $\mathcal{T}$ and the technology $\mathbf{M} \in \mathcal{M}_p$ defined in the same way as in Section 2, and $V_P(\mathcal{T}) = \inf_{\mathbf{M} \in \mathcal{M}_p} V_P(\mathcal{T}|M)$.

The problem that we studied in the paper restricted the planner to offering singleton menus of taxes. We show that tax progressivity remains optimal even if the planner has the ability to screen: Any menu of taxes is dominated by an alternative menu that contains exclusively convex taxes.

Proposition 6 (Menu of progressive taxes). *For every menu of tax rules $\mathcal{T}$, there exists a menu $\overline{\mathcal{T}}$ such that $|\overline{\mathcal{T}}| = \mathcal{T}$, T is convex for every $T \in \overline{\mathcal{T}}$, and $V_P(\overline{\mathcal{T}}) \geq V_P(\mathcal{T})$.*

Our next result establishes that, without loss of optimality, the planner can restrict attention to menus whose cardinality is at most the cardinality of the range of M_i^0. This result is not trivial, given that in principle the planner may use taxes to screen for the technology $\mathbf{M}$. It generalizes Theorem 4 in Carroll (2015), which states that screening does not help the principal whenever M_i^0 is degenerate. Let $\text{range}(\mathbf{M}^0) \equiv \{M \in \mathcal{M} : M_i^0 = M \text{ for some } i \in \mathcal{I}\}$ and suppose that $\text{range}(\mathbf{M}^0)$ is finite with cardinality equal to n—e.g., as in the case with finite types studied in Section 4.[35]

Proposition 7 (Upper bound on $|\mathcal{T}|$). *For every feasible menu of tax rules $\mathcal{T}$, there exists a feasible menu $\overline{\mathcal{T}} \subseteq \mathcal{T}$ such that $|\overline{\mathcal{T}}| \leq n$ and $V_P(\overline{\mathcal{T}}) \geq V_P(\mathcal{T})$.*

Providing a more general characterization of the optimal menu of taxes and identifying conditions under which the planner can strictly benefit from screening remains an interesting open question. A promising direction would be to explore conditions within the robust-contracting framework that enable the application of the "decoupling methods" developed by Castro-Pires et al. (2024), which simplify the analysis of problems that combine moral hazard and adverse selection.

[35] The result can be extended to a continuum of types if we further assume that $M^0 : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$ is continuous. In that case, without loss, the cardinality of an optimal menu has at most the cardinality of Θ.

Relationship between tax progressivity and inequality aversion. A common intuition for the use of progressive taxation in practice is that governments, potentially reflecting societal values, may strongly favor redistribution. This intuition is not supported by the canonical taxation model: there, regressive outcomes—such as zero marginal taxation of top incomes—can be optimal even under strong inequality aversion, including Rawlsian preferences. This discrepancy raises an important question: does a comparative static relationship between the degree of tax progressivity and the planner's preference for redistribution hold in our framework? Our analysis of the binary-types model suggests that the answer is inconclusive and depends on additional assumptions.

Extending this comparative static analysis to more general settings presents further conceptual challenges. Specifically, there is no universally accepted way to define an ordering of tax schedules based on progressivity. Throughout this paper, we have informally linked progressivity to the convexity of the tax schedule, but other factors also play a role. For instance, one could argue that the most progressive tax schedule involves full taxation combined with a lump-sum payment to all workers—a schedule that is affine, i.e., simultaneously convex and concave. To the best of our knowledge, such questions have not been systematically explored in the existing taxation literature, leaving scope for future research to better understand the conditions under which stronger preferences for redistribution lead to more progressive tax schedules.

Workers' risk aversion. The assumption that workers are risk-neutral is admittedly strong. It plays a role in our main result by enabling the sustenance of random income choices by workers, which, in turn, motivates the planner to hedge against worst-case income risk by offering a progressive tax. We highlight that, as a consequence of the uncertainty about the technology $\mathbf{M}$, the planner effectively faces uncertainty about workers' preferences over risk, since the cost associated with choosing each income lottery F is unknown to her. Because of this, there is a sense in which our result continues to hold if the social planner does not know workers' preferences over risk but entertains the possibility that workers might act as if they were risk-neutral.[36]

Alternatively, if we take the stance that the social planner *knows* workers' risk aversion, we can generalize our main result to the case where workers have a strictly concave Bernoulli utility over consumption, by establishing that the optimal *worker-welfare schedule* is progressive (in the sense of being concave in income). To that end, suppose that workers value consumption according to the continuous, strictly increasing, and concave utility function $\tilde{u} : \mathbb{R}_+ \to \mathbb{R}$. The other aspects of the model are as described in Section 2. Any tax rule $T(y)$ gives rise to a *utility contract* $u : Y \to \mathbb{R}$ given by $\tilde{u}(y - T(y))$. Since $\tilde{u}$ is invertible, any utility contract u uniquely pins down the associated tax rule T that implements it. Hence, the problem where the planner optimizes over feasible tax rules is equivalent to one where the planner optimizes over continuous utility contracts $u : Y \to \mathbb{R}$ subject to the constraint that $u(y) \geq \tilde{u}(0)$ for all $y \in Y$. Proposition 8 states that, if workers are risk-averse, then the analog of Theorem 1 holds in utility space.

Proposition 8 (Risk-aversion)**.** *There exists a worst-case optimal utility contract that is concave.*

[36] Walton and Carroll (2022) formalize this idea.

According to Proposition 8, there is an optimal tax rule that is progressive in the sense that the marginal *utility* that workers extract from their earned income decreases with their earnings. Given that it is workers' utility for consumption (not consumption itself) that the planner cares about, this notion is arguably a valid way to judge the progressivity of the tax system.

Alternatively, our model can be deemed as an approximation to settings in which workers have relatively low risk aversion. Under that interpretation, it is possible to extend our proof of Proposition 8 to show that, as workers' risk aversion vanishes, the optimum can be approximated arbitrarily well by a progressive tax rule. In particular, a commonly made assumption that has gained empirical support is that high-income earners have low risk aversion (Ogaki and Zhang, 2001). In light of this, our findings can be used to justify the use of progressive taxes, specifically at the top of the income distribution.

References

Akbarpour, M., P. Dworczak, and S. D. Kominers (2024): "Redistributive allocation mechanisms," Journal of Political Economy, 132, 1831–1875.

Aliprantis, C. D. and K. C. Border (2006): Infinite Dimensional Analysis, Springer.

Atkinson, A. B. (1995): Public economics in action: the basic income/flat tax proposal, Clarendon Press.

Aumann, R. J. and M. Maschler (1995): Repeated games with incomplete information, MIT press.

Barron, D., G. Georgiadis, and J. Swinkels (2020): "Optimal contracts with a risk-taking agent," Theoretical Economics, 15, 715–761.

Berliant, M. and M. Gouveia (2022): "On the Political Economy of Nonlinear Income Taxation," Working Paper.

Bhandari, A., J. Borovička, and Y. Yao (2024): "Robust bounds on optimal tax progressivity," Working Paper.

Boadway, R. and L. Jacquet (2008): "Optimal marginal and average income taxation under maximin," Journal of Economic Theory, 143, 425–441.

Boadway, R. and M. Sato (2014): "Optimal Income Taxation and Risk: The Extensive-Margin Case," Annals of Economics and Statistics/Annales d'Économie et de Statistique, 159–183.

Carroll, G. (2015): "Robustness and linear contracts," American Economic Review, 105, 536–563.

——— (2019): "Robustness in mechanism design and contracting," Annual Review of Economics, 11, 139–166.

Carroll, G. and D. Meng (2016): "Robust contracting with additive noise," Journal of Economic Theory, 166, 586–604.

Castro-Pires, H., H. Chade, and J. Swinkels (2024): "Disentangling moral hazard and adverse selection," American economic review, 114, 1–37.

Chassang, S. (2013): "Calibrated incentive contracts," Econometrica, 81, 1935–1971.

Chung, K.-S. and J. C. Ely (2007): "Foundations of dominant-strategy mechanisms," The Review of Economic Studies, 74, 447–476.

Diamond, P. (1998a): "Managerial incentives: On the near linearity of optimal compensation," Journal of Political Economy, 106, 931–957.

——— (1998b): "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," American Economic Review, 83–95.

Doval, L. and V. Skreta (2023): "Constrained information design," Mathematics of Operations Research.

Dworczak, P., S. D. Kominers, and M. Akbarpour (2021): "Redistribution through markets," Econometrica, 89, 1665–1698.

Farhi, E. and I. Werning (2013): "Insurance and taxation over the life cycle," Review of Economic Studies, 80, 596–635.

Fenchel, W. and D. W. Blackett (1953): Convex cones, sets, and functions, Princeton University, Department of Mathematics, Logistics Research Project.

Garrett, D. F. (2014): "Robustness of simple menus of contracts in cost-based procurement," Games and Economic Behavior, 87, 631–641.

Golosov, M., M. Troshkin, and A. Tsyvinski (2016): "Redistribution and social insurance," American Economic Review, 106, 359–386.

Hansen, L. P. and T. J. Sargent (2001): "Robust control and model uncertainty," American Economic Review, 91, 60–66.

Kambhampati, A. (2023): "Randomization is optimal in the robust principal-agent problem," Journal of Economic Theory, 207, 105585.

Kambhampati, A., J. Toikka, and R. Vohra (2024): "Randomization and the Robustness of Linear Contracts," Working Paper.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," American Economic Review, 101, 2590–2615.

Kang, Z. Y. (2023): "The Public Option and Optimal Redistribution," Working Paper.

Landier, A. and G. Plantin (2017): "Taxing the rich," The Review of Economic Studies, 84, 1186–1209.

Le Treust, M. and T. Tomala (2019): "Persuasion with limited communication capacity," Journal of Economic Theory, 184, 104940.

Lockwood, B. B., A. Sial, and M. Weinzierl (2021): "Designing, Not Checking, for Policy Robustness: An Example with Optimal Taxation," Tax Policy and the Economy, 35, 1–54.

Luenberger, D. G. (1997): Optimization by vector space methods, John Wiley & Sons.

Makris, M. and A. Pavan (2021): "Taxation under learning by doing," Journal of Political Economy, 129, 1878–1944.

Milgrom, P. and I. Segal (2002): "Envelope theorems for arbitrary choice sets," Econometrica, 70, 583–601.

Mirrlees, J. (1974): "Notes on welfare economics, information, and uncertainty. M. Balch, D. McFadden, S. Wu, eds. Essays on Economic Behavior Under Uncertainty," .

Mirrlees, J. A. (1971): "An exploration in the theory of optimum income taxation," The review of economic studies, 38, 175–208.

Ogaki, M. and Q. Zhang (2001): "Decreasing relative risk aversion and tests of risk sharing," Econometrica, 69, 515–526.

Pai, M. and P. Strack (2023): "Taxing Externalities Without Hurting the Poor," Available at SSRN 4180522.

Phelps, E. S. (1973): "Taxation of wage income for economic justice," The Quarterly Journal of Economics, 87, 331–354.

Piketty, T. and E. Saez (2013): "Optimal labor income taxation," in Handbook of public economics, Elsevier, vol. 5, 391–474.

Ramsey, F. P. (1927): "A Contribution to the Theory of Taxation," The economic journal, 37, 47–61.

Sadka, E. (1976): "On income distribution, incentive effects and optimal income taxation," The review of economic studies, 43, 261–267.

Saez, E. (2001): "Using elasticities to derive optimal income tax rates," The review of economic studies, 68, 205–229.

Saez, E. and S. Stantcheva (2016): "Generalized social marginal welfare weights for optimal tax theory," American Economic Review, 106, 24–45.

Seade, J. K. (1977): "On the shape of optimal tax schedules," Journal of public Economics, 7, 203–235.

Stantcheva, S. (2017): "Optimal taxation and human capital policies over the life cycle," Journal of Political Economy, 125, 1931–1990.

Tuomala, M. (2016): Optimal redistributive taxation, Oxford University Press.

Varian, H. R. (1980): "Redistributive taxation as social insurance," Journal of public Economics, 14, 49–68.

Walton, D. and G. Carroll (2022): "A general framework for robust contracting models," Econometrica, 90, 2129–2159.

Appendix

A Proofs

This section presents the proofs of all the main results in the paper. To enhance readability, the proofs of certain results, which involve technical details less critical to the core intuitions behind the results, are deferred to an Online Appendix.

A.1 Proof of Lemma 1

Let $\underline{V}(T,i) \equiv W(V_w(T|M_i^0),i) + \alpha\mathbb{E}_{F_i}[T(y)]$, with F_i defined as in Lemma 1. We seek to show that

$$V_P(T) = \int_{\mathcal{I}} \underline{V}(T,i)\, di.$$

We begin by showing that $V_P(T) \geq \int_{\mathcal{I}} \underline{V}(T,i)\, di$. To do so, we use the following claim. Let

$$r_i(T) = \min_{F \in \Delta(Y)} \mathbb{E}_F[T(y)], \quad \text{subject to } \mathbb{E}_F[y - T(y)] \geq V_w(T|M_i^0), \tag{A.1}$$

which is the value of the revenue-minimization problem defined in (3.2).

Claim 1.

$$W(V_w(T|M_i^0),i) + \alpha r_i(T) = \min_{(F,\phi)\in\Delta(Y)\times\mathbb{R}_+} \{W(\mathbb{E}_F[y-T(y)]-\phi),i) + \alpha\mathbb{E}_F[T(y)]\}, \tag{A.2}$$

$$\text{subject to } \mathbb{E}_F[y-T(y)] - \phi \geq V_w(T|M_i^0). \tag{A.3}$$

Moreover, there exists a solution to (A.2)-(A.3) in which (A.3) holds with equality.

Proof. See Online Appendix OA.2.1. □

Fix a plausible technology $\mathbf{M} \in \mathcal{M}_p$. Let $(\tilde{F}_i, \tilde{\phi}_i)$ be worker i's choice under M_i and let F_i be defined as in Lemma 1. First, if i is such that $V_w(T|M_i^0) = \max_{y\in Y}\{y - T(y)\}$, then i can secure maximal consumption at zero cost under M_i^0. Therefore, under any $M_i \supseteq M_i^0$, i's choice $\tilde{F}_i$ will be supported on a subset of $Y^* \equiv \arg\max_{y\in Y}\ \{y - T(y)\}$ and $V_w(T|M_i^0) = V_w(T|M_i)$. Since the worker breaks ties in favor of the social planner and $M_i \supseteq M_i^0$ it follows that $\mathbb{E}_{\tilde{F}_i}[T(y)] \geq \mathbb{E}_{F_i}[T(y)]$. Thus, $W(V_w(T|M_i),i) + \alpha\mathbb{E}_{\tilde{F}_i}[T(y)] \geq \underline{V}(T,i)$. Second, consider i such that $V_w(T|M_i^0) < \max_{y\in Y}\{y - T(y)\}$. Since $M_i \supseteq M_i^0$, it holds that $(\tilde{F}_i, \tilde{\phi}_i)$ satisfies (A.3). Thus, by Claim 1,

$$W(V_w(T|M_i),i) + \alpha\mathbb{E}_{\tilde{F}_i}[T(y)] \geq W(V_w(T|M_i^0),i) + \alpha r_i(T) = \underline{V}(T,i).$$

Combining these two points and integrating across i we have

$$V_P(T|\mathbf{M}) \geq \int_{\mathcal{I}} \underline{V}(T,i)\, di = \int_{\mathcal{I}} \{W(V_w(T|M_i^0),i) + \alpha\mathbb{E}_{F_i}[T(y)]\}\, di.$$

Since $\mathbf{M}$ was arbitrary, the inequality holds if we substitute $V_P(T|\mathbf{M})$ for $V_P(T)$.

Next, we show that $V_P(T) \leq \int_{\mathcal{I}} \underline{V}(T,i)\, di$. Let $y^* \equiv \arg\max_{y\in Y^*}\ T(y)$. For $\varepsilon \in [0,1]$, let $(F_i^\varepsilon, \phi_i^\varepsilon) = ((1-\varepsilon)F_i + \varepsilon\delta_{y^*}, (1-\varepsilon)(\mathbb{E}_{F_i}[y - T(y)] - V_w(T|M_i^0)))$, and define the technology

$$M_i^\varepsilon \equiv M_i^0 \cup \{(F_i^\varepsilon, \phi_i^\varepsilon)\}.$$

By the Measurable Maximum Theorem (Theorem 17.18 in Aliprantis and Border (2006)), M_i^ε is measurable and thus $\mathbf{M}^\varepsilon \equiv (M_i^\varepsilon)_{i\in\mathcal{I}}$ is plausible. We show that total welfare under $\mathbf{M}^\varepsilon$ approaches $\int_{\mathcal{I}} \underline{V}(T,i)\, di$ as $\varepsilon \to 0$.

Claim 2. For any feasible T, $\lim_{\varepsilon\downarrow 0} V_P(T|\mathbf{M}^\varepsilon) = \int_{\mathcal{I}} \underline{V}(T,i)\,di$.

Proof. Consider first i such that $V_w(T|M_i^0) < y^* - T(y^*)$. Worker i's uniquely optimal choice under M_i^ε is to choose $(F_i^\varepsilon, \phi_i^\varepsilon)$, which gives him a payoff of $(1-\varepsilon)V_w(T|M_i^0) + \varepsilon(y^* - T(y^*)) > V_w(T|M_i^0)$. On the other hand, if $V_w(T|M_i^0) = y^* - T(y^*)$, then as argued before, worker i's choice takes the form $(F, 0)$ with F fully supported on Y^*. By the definition of $(F_i^\varepsilon, \phi_i^\varepsilon)$ and y^*, it holds that $\mathbb{E}_{F_i^\varepsilon}[y - T(y)] - \phi_i^\varepsilon = V_w(T|M_i^0)$ and that $\mathbb{E}_{F_i^\varepsilon}[T(y)] \geq \mathbb{E}_{F_i}[T(y)]$. Thus $(F_i^\varepsilon, \phi_i^\varepsilon)$ is an optimal choice for worker i (after applying favorable tie-breaking) under M_i^ε.

As a consequence, we have

$$\lim_{\varepsilon\downarrow 0} V_P(T|\mathbf{M}^\varepsilon) =$$
$$\lim_{\varepsilon\downarrow 0} \int_{\mathcal{I}} \{W((1-\varepsilon)V_w(T|M_0^i) + \varepsilon(y^* - T(y^*)), i) + \alpha[(1-\varepsilon)\mathbb{E}_{F_i}[T(y)] + \varepsilon T(y^*)]\}\,di.$$

For all $\varepsilon \in (0,1)$ and $i \in \mathcal{I}$, it holds that

$$\begin{aligned} &|W((1-\varepsilon)V_w(T|M_0^i) + \varepsilon(y^* - T(y^*)), i) + \alpha[(1-\varepsilon)\mathbb{E}_{F_i}[T(y)] + \varepsilon T(y^*)]| \\ &\qquad \leq \sup_{i\in\mathcal{I}} |W(y^* - T(y^*), i)| + \alpha \max_{y\in Y} |T(y)| < +\infty. \end{aligned}$$

Thus, by the Dominated Convergence Theorem,

$$\begin{aligned} \lim_{\varepsilon\downarrow 0} V_P(T|\mathbf{M}^\varepsilon) &= \int_{\mathcal{I}} \lim_{\varepsilon\downarrow 0} \{W((1-\varepsilon)V_w(T|M_0^i) + \varepsilon(y^* - T(y^*)), i) + \alpha[(1-\varepsilon)\mathbb{E}_{F_i}[T(y)] + \varepsilon T(y^*)]\}\,di \\ &= \int_{\mathcal{I}} \underline{V}(T,i)\,di. \end{aligned}$$

□

Applying Claim 2, we have $V_P(T) \leq \lim_{\varepsilon\downarrow 0} V_P(T|\mathbf{M}^\varepsilon) = \int_{\mathcal{I}} \underline{V}(T,i)\,di$.

A.2 Proof of Theorem 1

Let T be a feasible tax rule and write $c(y) = y - T(y)$. We begin by showing that, for every feasible T, there exists an alternative feasible $\overline{T}$, which is double-monotone and convex, such that $V_P(T) \leq V_P(\overline{T})$ (Steps 1 and 2). Then, we show that a worst-case optimal tax exists (Step 3).

Step 1: i-specific affine rules. We begin by constructing an affine tax rule for every i that dominates the original one in terms of worst-case welfare coming from worker i.

Fix $i \in \mathcal{I}$ and let F_i be defined as in Lemma 1. Let

$$A \equiv \mathrm{co}\{(c(y), y - c(y)) \in \mathbb{R}^2 : y \in Y\}, \quad B_i \equiv \{(u,v) \in \mathbb{R}^2 : u > V_w(T|M_i^0), v < E_{F_i}[T(y)]\},$$

where $\mathrm{co}(A)$ stands for the convex hull of the set A. The definition of F_i implies that A and B_i are disjoint. Since the two sets are convex, by the Separating Hyperplane Theorem, there exists $(\kappa_i, \beta_i, \lambda_i) \in \mathbb{R}^3$ with $(\beta_i, \lambda_i) \neq 0$ such that

$$\kappa_i + \lambda_i c(y) - \beta_i(y - c(y)) \leq 0, \quad \forall y \in Y \tag{A.4}$$
$$\kappa_i + \lambda_i u - \beta_i v \geq 0, \quad \forall (u,v) \in B_i. \tag{A.5}$$

Since $(\mathbb{E}_{F_i}[c(y)], \mathbb{E}_{F_i}[T(y)])$ belongs to the closure of B_i, we also have

$$\kappa_i + \lambda_i \mathbb{E}_{F_i}[c(y)] - \beta_i \mathbb{E}_{F_i}[T(y)] = 0. \tag{A.6}$$

By (A.5) and the definition of B_i, we have that $\lambda_i \geq 0$ and $\beta_i \geq 0$. By (A.4), $\kappa_i \leq -(\lambda_i + \beta_i)c(0) \leq 0$. Moreover, if $\beta_i = 0$, then (A.4) and (A.5) imply that $V_w(T|M_i^0) \geq \max_{y\in Y} c(y)$.

Additionally, we show that, if $\beta_i > 0$, then

$$\mathbb{E}_{F_i}[T(y)] = \frac{\kappa_i + \lambda_i V_w(T|M_i^0)}{\beta_i}. \tag{A.7}$$

If $\lambda_i = 0$, then (A.7) follows immediately from (A.6). If $\lambda_i > 0$ and $\beta_i > 0$, then (A.5) implies that $\kappa_i + \lambda_i V_w(T|M_i^0) - \beta_i \mathbb{E}_{F_i}[T(y)] \geq 0$; while on the other hand we have that $\mathbb{E}_{F_i}[c(y)] \geq V_w(T|M_i^0)$ and $\kappa_i + \lambda_i \mathbb{E}_{F_i}[c(y)] - \beta_i \mathbb{E}_{F_i}[T(y)] = 0$. The two assertions together with $\lambda_i > 0$ imply that $V_w(T|M_i^0) = \mathbb{E}_{F_i}[c(y)]$ and thus (A.7) holds.

For each $i \in \mathcal{I}$, we define the following non-negative affine consumption rule

$$c_i(y) = \frac{\beta_i y - \kappa_i}{\beta_i + \lambda_i}. \tag{A.8}$$

Step 2: Constructing a dominating tax rule. We now use the collection $\{c_i\}_{i\in\mathcal{I}}$ to construct a consumption rule that outperforms c. Let

$$\bar{c}(y) = \inf_{i\in\mathcal{I}} c_i(y), \tag{A.9}$$

and $\overline{T}(y) = y - \bar{c}(y)$. $\overline{T}$ satisfies the properties of Theorem 1.[37] By (A.4), $c_i(y) \geq c(y)$ for all $i \in \mathcal{I}$ and $y \in Y$, and hence $\bar{c}(y) \geq c(y)$. It follows that $V_w(\overline{T}|M_i^0) \geq V_w(T|M_i^0)$ for all $i \in \mathcal{I}$, and thus worst-case worker-welfare is higher under $\overline{T}$.

We now show that worst-case revenue is also higher under $\overline{T}$ for every $i \in \mathcal{I}$. To that end, consider first $i \in \mathcal{I}$ such that $\beta_i > 0$. For any $\tilde{F} \in \Delta(Y)$ satisfying $\mathbb{E}_{\tilde{F}}[\bar{c}(y)] \geq V_w(\overline{T}|M_i^0)$ (by Lemma 1 it suffices to restrict attention to income choices satisfying this condition), we can write

$$\mathbb{E}_{\tilde{F}}[y - \bar{c}(y)] \geq \mathbb{E}_{\tilde{F}}[y - c_i(y)] = \frac{\kappa_i + \lambda_i \mathbb{E}_{\tilde{F}}[c_i(y)]}{\beta_i} \geq \frac{\kappa_i + \lambda_i V_w(\overline{T}|M_i^0)}{\beta_i} \tag{A.10}$$

$$\geq \frac{\kappa_i + \lambda_i V_w(T|M_i^0)}{\beta_i} = \mathbb{E}_{F_i}[T(y)], \tag{A.11}$$

where the final equality follows from (A.7).

Second, suppose that $\beta_i = 0$, and thus $\mathbb{E}_{F_i}[c(y)] = V_w(T|M_i^0) = -\kappa_i/\lambda_i$. Since $c(y) \leq c_i(y) = -\kappa_i/\lambda_i$ for all y, it follows that $c(y) = -\kappa_i/\lambda_i$ for all y in the support of F_i. Further, since $c_{i'}(y) \geq c(y)$ for all $i' \in \mathcal{I}$ and $y \in Y$, and equality holds under $i' = i$ and y in the support of F_i, we have that $\bar{c}(y) = -\kappa_i/\lambda_i = c(y)$ for all y in the support of F_i. Thus, under $\overline{T}$ and any $M_i \supseteq M_i^0$, worker i's payoff is maximized by choosing $(F_i, 0) \in M_i^0$, and this choice yields revenue equal to $\mathbb{E}_{F_i}[T(y)] = \mathbb{E}_{F_i}[\overline{T}(y)]$. Given that the worker breaks ties in favor of the planner, the revenue from worker i under any M_i and $\overline{T}$ is at least $\mathbb{E}_{F_i}[T(y)]$, that is worst-case revenue under T.

In sum, for each $i \in \mathcal{I}$, $\overline{T}$ yields a worst-case improvement over T in terms of both workers' welfare and revenue, and thus $V_P(\overline{T}) \geq V_P(T)$.

Step 3: Existence. The above steps establish that, for every feasible T, there is a $\overline{T}$ which is *(i)* double-monotone and *(ii)* convex, that satisfies $V_P(\overline{T}) \geq V_P(T)$. We further show in Lemma 3 that $y - T(y)$ is without loss bounded above. To that end, let

$$\overline{\phi} \equiv \sup_{i\in\mathcal{I}} \{\phi_i : (\phi_i, F) \in M_i^0\},$$

and $\overline{C} \equiv C + \overline{\phi}$, where C is defined as in Assumption 2.

Lemma 3. *For any feasible T such that $c(y) = y - T(y)$ is non-decreasing, there exists a tax rule $\hat{T}$ satisfying $\hat{T}(0) \geq -\overline{C}$ such that $V_P(\hat{T}) \geq V_P(T)$,*

[37] Convexity of $\overline{T}$ follows from the fact that $\bar{c}$, as the pointwise infimum of a family of affine functions, is concave.

Proof. See Online Appendix OA.2.2. □

The proof is completed by the following lemma showing that the maximum exists.

Lemma 4. *There exists a worst-case optimal tax rule.*

Proof. Let $\hat{\mathbf{T}} \subset \mathbf{T}_f$ be the set of continuous, convex, and double-monotone functions from Y to $[-\overline{C}, \overline{y}]$. The arguments above, together with Lemma 3, imply that it suffices to show that $\max_{T\in\hat{\mathbf{T}}} V_P(T)$ is well-defined. The family $\hat{\mathbf{T}}$ is uniformly bounded and equicontinuous[38], and is thus compact by the Arzelà-Ascoli Theorem.

Claim 3. $V_P(T)$ is upper semi-continuous on $\hat{\mathbf{T}}$.

Proof. See Online Appendix OA.2.3. □

Existence of the maximum follows from the fact that the objective is upper semi-continuous and the choice set compact. □

A.3 Proof of Theorem 1*

Take any feasible $T \in \mathbf{T}_f$ and let $c(y) = y - T(y)$. Let $\tilde{c}(y)$ be the concavifiction of $c(y)$ and $\tilde{T}(y) = y - \tilde{c}(y)$. Applying Claim 8 below (Section A.11) to the singleton menus of taxes $\mathcal{T} = \{T\}$ and $\tilde{\mathcal{T}} = \{\tilde{T}\}$, we have that for all $\mathbf{M} \in \mathcal{M}_p$, there exists $\mathbf{M}' \in \mathcal{M}_p$ such that $V_P(T|\mathbf{M}') = V_P(\tilde{T}|\mathbf{M})$. Thus,

$$V_P(\tilde{T}|\mathbf{M}) = V_P(T|\mathbf{M}') \geq V_P(T).$$

Since $\mathbf{M}$ was arbitrary, it follows that $V_P(\tilde{T}) \geq V_P(T)$ as desired.

A.4 Proof of Proposition 1

Let $T(y)$ be a worst-case optimal tax rule and $c(y) = y - T(y)$ be the associated consumption rule. Set $\bar{c}(y)$ to be the consumption rule defined in (A.9), and $\tilde{c}(y)$ to be the concavification of c, and write $\overline{T}(y) = y - \bar{c}(y)$ and $\tilde{T}(y) = y - \tilde{c}(y)$. We make use of the following claim.

Claim 4. $c(y) \leq \tilde{c}(y) \leq \bar{c}(y)$.

Proof. Recall that $\tilde{c}(y)$ is defined as the smallest concave function that is pointwise higher than c. The first inequality follows immediately from this definition of $\tilde{c}(y)$. Moreover, $\bar{c}$ is itself concave and majorizes c. Therefore, $\bar{c}(y) \geq \tilde{c}(y)$ follows again from the definition of $\tilde{c}$. □

Let $Y^* = \arg\max_{y\in Y} c(y)$ and $y^* = \arg\max_{y\in Y^*} T(y)$. Take any $F_i^0 \in \mathcal{F}_i^0(T)$ and $\underline{F}_i(T) \in \underline{\mathcal{F}}_i(T)$ for each $i \in \mathcal{I}$, and let $F^0 \in \mathcal{F}^0(T)$ and $\underline{F} \in \underline{\mathcal{F}}(T)$ be the associated aggregate distributions. Set $\phi_i^0 \in \mathbb{R}_+$ to be such that $(F_i^0, \phi_i^0) \in X_w(T|M_i^0)$. Let $\mathcal{J} \equiv \{i \in \mathcal{I} : V_w(T|M_i^0) < c(y^*)\}$, which is measurable under our assumptions. For all $i \in \mathcal{I} \setminus \mathcal{J}$, it holds that F_i^0 is supported on a subset of y^* and that $\mathcal{F}_i^0(T) = \underline{\mathcal{F}}_i(T)$ (by Lemma 1). Since $c(y) = \bar{c}(y)$ for all y the support of F_i^0 (shown in Step 2 of the proof of Theorem 1), it follows from Claim 4 that $c(y) = \tilde{c}(y) = \bar{c}(y)$ F_i^0- and $\underline{F}_i$-almost everywhere for any $i \in \mathcal{I} \setminus \mathcal{J}$.

[38] To prove equicontinuity, we argue that all the functions in $\hat{\mathbf{T}}$ have a common Lipschitz constant given by $K = 1$. To that end, let $T \in \hat{\mathbf{T}}$ and $y_1, y_2 \in (0, \bar{y})$. Convexity of T implies that we can write $T(y_1) - T(y_2) = \int_{y_2}^{y_1} T'(y)\,dy$, where T' is the right-derivative of T. The fact that $y - T(y)$ is non-decreasing then implies that $T'(y) \leq 1$ and thus $|T(y_1) - T(y_2)| \leq |y_1 - y_2|$. The argument for the case in which $y_i \in \{0, \bar{y}\}$ for some $i = 1, 2$ follows from the previous step and continuity of T.

Next, consider $i \in \mathcal{J}$. As we showed in Step 2 of Theorem 1, $\beta_i > 0$ for all $i \in \mathcal{J}$. We then have

$$V_P(\overline{T}) - V_P(T) \geq \int_{\mathcal{J}} \Bigg(W(V_w(\overline{T}|M_i^0), i) - W(V_w(T|M_i^0), i)] + \\ \alpha \frac{\lambda_i[V_w(\overline{T}|M_i^0) - V_w(T|M_i^0)]}{\beta_i} \Bigg)\, di \geq \int_{\mathcal{J}} [W(\mathbb{E}_{F_i^0}[\bar{c}(y)] - \phi_i^0, i) - W(\mathbb{E}_{F_i^0}[c(y)] - \phi_i^0, i)]\, di,$$

where the first inequality follows from (A.10) and (A.11) and the fact that $\overline{T}$ yields a weak welfare improvement conditional on each $i \in \mathcal{I}$, and the second inequality uses the fact that $V_w(\overline{T}|M_i^0) - V_w(T|M_i^0) \geq 0$ and $V_w(\overline{T}|M_i^0) \geq \mathbb{E}^0_{F_i}[\bar{c}(y)] - \phi_i^0$ for all $i \in \mathcal{I}$. If $W(\cdot, i)$ is strictly increasing for all $i \in \mathcal{I}$, optimality of T requires that $\mathbb{E}_{F_i^0}[\bar{c}(y)] = \mathbb{E}_{F_i^0}[c(y)]$ almost everywhere in $i \in \mathcal{J}$. Combining everything and applying Claim 4, we obtain that $\bar{c}(y) = \tilde{c}(y) = c(y)$ F^0-almost everywhere.

It remains to show that $\int_{\mathcal{J}} \int_Y [\bar{c}(y) - c(y)]\, d\underline{F}_i(y)\, di = 0$. We know from optimality of T and the previous paragraph that for all $i \in \mathcal{J}$, $V_w(\overline{T}|M_i^0) = V_w(T|M_i^0) < \max_{y \in Y} c(y) \leq \max_{y \in Y} \bar{c}(y)$. Moreover, optimality of T requires that $\int_{\mathcal{J}} [r_i(\overline{T}) - r_i(T)]\, di = 0$ with $r_i(T)$ as defined in (A.1). Thus,

$$\int_{\mathcal{J}} r_i(\overline{T})\, di \leq \int_{\mathcal{J}} \mathbb{E}_{\underline{F}_i}[\overline{T}(y)]\, di \leq \int_{\mathcal{J}} \mathbb{E}_{\underline{F}_i}[T(y)]\, di = \int_{\mathcal{J}} r_i(T)\, di = \int_{\mathcal{J}} r_i(\overline{T})\, di,$$

where the first inequality follows from the definition of $r_i(\overline{T})$ and the fact that $\mathbb{E}_{\underline{F}_i}[y - \overline{T}(y)] \geq \mathbb{E}_{\underline{F}_i}[y - T(y)] \geq V_w(T|M_i^0) = V_w(\overline{T}|M_i^0)$ almost everywhere in $i \in \mathcal{J}$, and the second inequality from the fact that $\overline{T}(y) \leq T(y)$ for all $y \in Y$. Hence, it must be that $\mathbb{E}_{\underline{F}_i}[\overline{T}(y)] = \mathbb{E}_{\underline{F}_i}[T(y)]$ almost everywhere in $i \in \mathcal{J}$, which together with Claim 4 yields the result.

A.5 Proof of Proposition 2

Existence of a worst-case optimal tax within the affine class follows from a straightforward generalization of the arguments in Lemma 6 of Carroll (2015). We omit this step for brevity.

Let the tax rule $T_a(y) = t^* + \tau^* y$ be worst-case optimal within the affine class. Suppose that $\tau^* < 1$. For T_a to be fully worst-case optimal, it has to be the case that T_a is not strictly dominated by any feasible perturbation of the form

$$T^\varepsilon(y) \equiv T(y) + \varepsilon D(y),$$

where $\varepsilon \in \mathbb{R}$ and $D(y)$ is a continuous function.

In Appendix B, we show that $V_P(T^\varepsilon)$ is directionally differentiable at $\varepsilon = 0$ and provide an expression for its derivatives. Here, we consider the perturbation with $D(y) = -\min\{\tilde{y} - y, 0\}$ and $\tilde{y} > 0$. Observe that $T^\varepsilon(y)$ is strictly progressive. The fact that $\tau^* < 1$ and $\tilde{y} > 0$ ensures that non-negativity of consumption is still satisfied under T^ε for $\varepsilon > 0$ sufficiently small, and thus T^ε is feasible. Claim 5 describes the effect of this perturbation.

Claim 5. If $T(y) = t + \tau y$ with $t \leq 0$ and $\tau \in [0, 1)$, and if $D(y) = -\min\{\tilde{y} - y, 0\}$ for some $\tilde{y} \in (0, \overline{y}]$, then

$$\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\bigg|_{\varepsilon = 0} = \\ \int_{\mathcal{I}} \left[(w_i - \alpha) \max_{F \in \mathcal{F}_i^0} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] + \alpha \frac{\max_{F \in \mathcal{F}_i^0} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] - \min\{\tilde{y} - y_i, 0\}}{1 - \tau} \right] di.$$

Proof. See Online Appendix OA.2.4. □

We then have the following necessary condition for worst-case optimality of T_a:

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} \leq 0.$$

By Claim 5, the derivative is bounded below by

$$\begin{aligned}\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} &\geq \int_{\mathcal{I}} \left[\left(w_i + \alpha\frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F_i^0}[\min\{\tilde{y}-y,0\}] - \alpha\frac{\min\{\tilde{y}-y_i^0+\phi_i^0/(1-\tau^*),0\}}{1-\tau^*}\right] di \\ &\geq \int_{\mathcal{I}} \left[\left(w_i + \alpha\frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F_i^0}[\min\{\tilde{y}-y,0\}] - \alpha\mathbf{I}(y_i^0 > \tilde{y})\frac{\tilde{y}-y_i^0+\phi_i^0/(1-\tau^*)}{1-\tau^*}\right] di.\end{aligned}$$

Then, if condition (3.4) holds, $\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} > 0$, contradicting worst-case optimality of T_a.

A.6 Proof of Proposition 3

Let $T_a(y) = t^* + \tau^* y$ be worst-case optimal within the affine class. We begin by computing the planner's payoff from any affine tax $T(y) = t + \tau y$. We slightly abuse notation by writing $(y_i^0, \phi_i^0) \in M_i^0$ when there exists $(F_i^0, \phi_i^0) \in M_i^0$ such that $\mathbb{E}_{F_i^0}[y] = y_i^0$.

Claim 6. If $T(y) = t + \tau y$ with $\tau \in [0,1]$, then

$$V_P(T) = \int_{\mathcal{I}} \left[W(-t + (1-\tau)y_i^0(\tau) - \phi_i^0(\tau), i) + \alpha\frac{\tau[-t+(1-\tau)y_i^0(\tau) - \phi_i^0(\tau)] + t}{1-\tau}\right] di,$$

with the interpretation that if $\tau = 1$, then

$$\frac{\tau[-t+(1-\tau)y_i^0 - \phi_i^0] + t}{1-\tau} = \begin{cases} t + y_i^0, & \text{if } \phi_i^0 = 0, \\ t, & \text{if } \phi_i^0 > 0. \end{cases}$$

Proof. See Online Appendix OA.2.5. □

Applying Claim 6, a necessary first-order condition for an optimal (t, τ) is

$$\int_{\mathcal{I}} [w_i(t,\tau) - \alpha]\, di \leq 0, \quad \text{with equality if } t < 0. \tag{A.12}$$

Assumption 2 ensures that, for every $\tau \in [0,1]$, there exists $t(\tau) \in \mathbb{R}_-$ satisfying (A.12). Also, whenever $\tau < 1$ we obtain the following additional necessary condition,[39]

$$\int_{\mathcal{I}} \left[(w_i(t(\tau),\tau) - \alpha)y_i^0(\tau) + \alpha\frac{\phi_i^0(\tau)}{(1-\tau)^2}\right] di \geq 0, \text{ if } \tau \in [0,1), \text{ with equality if } \tau > 0. \tag{A.13}$$

Rearranging terms in (A.13) we obtain (3.6).

Suppose that there exists τ^o satisfying the first-order condition (A.13). Because the objective has a discontinuity at $\tau = 1$, we need to verify that τ^o is not dominated by choosing $\tau = 1$. Thus, $\tau^0 < 1$ is optimal only if (3.7) is satisfied. Otherwise, if there is no $\tau^o \in [0,1)$ satisfying (3.6) and (3.7), then it must be that $\tau^* = 1$ (recall that existence of an optimal pair (t^*, τ^*) was established in Proposition 2).

[39] This expression uses the fact that, by the envelope theorem (Milgrom and Segal, 2002), $\frac{\partial V_w(T|M_i^0)}{\partial \tau} = y_i^0(\tau)$.

A.7 Proof of Lemma 2

(i) Suppose by contradiction that F_0^j is not a degenerate lottery. By Assumption 4, there exists $\phi' < \phi_0^j$ such that $(\delta_{y_0^j}, \phi') \in M^0(\theta)$. We then have

$$\mathbb{E}_{F_0^j}[y - T(y)] - \phi_0^j \leq y_0^j - \phi_0^j < y_0^j - \phi',$$

where the first inequality is a consequence of Jensen's inequality and convexity of $T(y)$. This contradicts optimality of (F_0^j, ϕ_0^j) for the worker.

(ii) Suppose by contradiction that $j > j'$ and $y_0^j < y_0^{j'}$. Incentive compatibility of types θ^j and $\theta^{j'}$ implies that

$$\Phi(y_0^{j'}, \theta^j) + \Phi(y_0^j, \theta^{j'}) \geq \Phi(y_0^{j'}, \theta^{j'}) + \Phi(y_0^j, \theta^j). \tag{A.14}$$

On the other hand, if $y_0^j < y_0^{j'}$, Assumption 5 implies that

$$\Phi(y_0^{j'}, \theta^j) + \Phi(y_0^j, \theta^{j'}) \leq \Phi(y_0^{j'}, \theta^{j'}) + \Phi(y_0^j, \theta^j).$$

Further, since y_0^j is feasible for type θ^j and $y_0^{j'}$ is feasible for type $\theta^{j'}$, the right-hand-side of this inequality is finite and therefore, $\Phi(y_0^{j'}, \theta^j) + \Phi(y_0^j, \theta^{j'}) < +\infty$. Applying Assumption 5 again, we thus have

$$\Phi(y_0^{j'}, \theta^j) + \Phi(y_0^j, \theta^{j'}) < \Phi(y_0^{j'}, \theta^{j'}) + \Phi(y_0^j, \theta^j),$$

which stands in contradiction with (A.14).

Additionally, if $y_0^j = 0$, then Assumption 5 implies that $0 = V_w(T|M^0(\theta^j)) = V_w(T|M^0(\theta^{j'}))$. Suppose now that $y_0^j > 0$. Applying Assumption 5 evaluated at $y = 0$ and $y' \in \{y_0^j, y_0^{j'}\}$, we have that $\Phi(y', \theta) \leq \Phi(y', \theta')$ and the inequality is strict if $y' > 0$. In particular, $y_0^{j'}$ is feasible for type θ^j. Therefore,

$$V_w(T|M^0(\theta^j)) = c(y_0^j) - \Phi(y_0^j, \theta) \geq c(y_0^{j'}) - \Phi(y_0^{j'}, \theta^j) \geq c(y_0^{j'}) - \Phi(y_0^{j'}, \theta^{j'}) = V_w(T|M^0(\theta j')),$$

and the second inequality is strict if $y_0^{j'} > 0$.

A.8 Proof of Proposition 4

Let T be a worst-case optimal tax schedule (which we know exists from Theorem 1) and let $\overline{T}$ be defined by (A.9). By definition, under the assumptions of Section 4, $\overline{T}$ is at-most n-piece affine. As we showed in Theorem 1, $V_P(\overline{T}) \geq V_P(T)$. Thus, the social planner's problem can be reduced to choosing $(k, \hat{\mathbf{y}}, t, \boldsymbol{\tau})$ to maximize worst-case welfare and using the tax schedule $T^k(y; \hat{\mathbf{y}}, t, \boldsymbol{\tau})$.

We also know from Theorem 1 that we can restrict attention to $t \in [-\overline{C}, 0]$ for some constant $\overline{C} \in \mathbb{R}_+$ (see Lemma 3) and to $\boldsymbol{\tau} \in [0,1]^k$ (by double-monotonicity). The convexity property in the theorem implies that we can restrict attention to $\tau^1 \leq ... \leq \tau^k$. As a result, for each $k \in \{1, ..., n\}$, the set of tax schedules of the form of (4.1) that satisfy the conditions in Theorem 1 is isomorphic to a compact subset of $Y^{k-1} \times [-\overline{C}, 0] \times [0,1]^k$ which is compact. By standard compactness and upper semi-continuity arguments, it follows that there exists $k \in \{1, ..., n\}$ and a worst-case optimal tax that is convex, double-monotone, and k-piece affine.

Moreover, starting from a worst-case optimal tax schedule T^k that is k-piece affine, it is possible to construct a simple tax schedule T^j with $j \leq k$ that satisfies $V_P(T^j) \geq V_P(T^k)$: This is obtained by iteratively applying the procedure described by (A.9) until the resulting schedule is simple.[40] Thus, T^j is a simple tax schedule that is worst-case optimal.

[40] The process always concludes after at most $k - 1$ steps, given that any affine tax schedule is simple.

A.9 Proof of Proposition 5

We make use of the following lemma describing the planner's objective under a k-piece affine tax.

Lemma 5. *Any worst-case optimal tax rule* $T^k(\cdot;\hat{\mathbf{y}},t,\boldsymbol{\tau})$ *must solve*

$$\max_{\mathbf{y_0}} V(\mathbf{y}_0), \tag{A.15}$$

where

$$V(\mathbf{y}_0) \equiv \max_{\hat{\mathbf{y}},t,\boldsymbol{\tau},\mathbf{y}} \sum_{j=1}^{n} \left(\int_{\mathcal{I}} \mathbb{I}(\theta_i = \theta^j) W(c^k(y_0^j;\hat{\mathbf{y}},t,\boldsymbol{\tau}) - \Phi(y_0^j,\theta^j),i)\, di + \alpha p^j T^k(y^j;\hat{\mathbf{y}},t,\boldsymbol{\tau}) \right) \tag{Opt$_{\text{fin}}$}$$

$$\textit{subject to } y^j = \begin{cases} y_0^j, & \textit{if } c^k(y_0^j;\hat{\mathbf{y}},t,\boldsymbol{\tau}) - \Phi(y_0^j,\theta^j) = c^k(\bar{y};\hat{\mathbf{y}},t,\boldsymbol{\tau}); \\ \min\{y \in Y : c^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau}) \geq c^k(y_0^j;\hat{\mathbf{y}},t,\boldsymbol{\tau}) - \Phi(y_0^j,\theta^j)\}, & \textit{if otherwise;} \end{cases} \tag{A.16}$$

$$\hat{y}^1 \leq \ldots \leq \hat{y}^{k-1}, t \leq 0, 0 \leq \tau^1 \leq \ldots \leq \tau^k \leq 1. \tag{A.17}$$

Proof. See Online Appendix OA.2.6. □

Using Lemma 5, we can derive necessary conditions for worst-case optimality of the *simple* tax schedule $T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau})$ for a fixed choice of $\mathbf{y}_0$. First, it cannot be that there is a strict welfare improvement from slightly perturbing the value of t. Because changing t does not affect worst-case income choices as defined in (A.16), we obtain the following necessary condition for optimality:

$$\sum_{j=1}^{n} p^j \omega^j - \alpha \leq 0, \text{ with equality if } t < 0.$$

Similarly, we derive necessary conditions for $\boldsymbol{\tau}$ by differentiating the objective in (Opt$_{\text{fin}}$). First, if $\tau^k < 1$, the condition ensuring that it is not strictly beneficial to perturb τ^k by a small amount is

$$\sum_{j=1}^{n} p^j \left(\mathbb{I}(y_0^j \geq \hat{y}^{k-1})(\alpha - \omega^j)(y_0^j - \hat{y}^{k-1}) - \frac{\alpha}{(1-\tau^k)^2} \mathbb{I}(y^j \geq \hat{y}^{k-1}) \Phi(y_0^j,\theta^j) \right) \leq 0, \text{ with equality if } \tau^k > 0. \tag{A.18}$$

We now rule out the possibility that $\tau^k = 1$ is optimal. It follows from equation (A.18) that $\tau^k = 1$ being optimal would imply that $\Phi(y_0^j,\theta^j) = 0$ for all $j \in \{1,...,n\}$ satisfying $y^j \geq \hat{y}^{k-1}$. If not, it would be strictly beneficial to slightly reduce τ^k—note that the fact that T^k is simple implies that $y^j \geq \hat{y}^{k-1}$ for some $j \in \{1,...,n\}$. Then, by part (i) of Assumption 5 and the definition of a simple tax schedule, it follows that $\hat{y}^{k-1} = y_0^j = y^j = 0$ for all $j \in \{1,...,n\}$—i.e., $T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau})$ is affine with slope equal to 1. Then, worst-case welfare under $T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau})$ is equal to $\int_{\mathcal{I}} W(-t,i)\, di + \alpha t$.

Consider using the the alternative tax schedule $T(y) = t + \varepsilon y$ where $\varepsilon > 0$ is arbitrarily small. By part (ii) of Assumption 5, there is a strictly positive measure of types that would choose to produce $y^j > 0$ under any technology. Thus, worst-case welfare under the new tax schedule is

$$\sum_{j=1}^{n} \int_{\mathcal{I}} \mathbb{I}(\theta_i = \theta^j)[W(-t + (1-\varepsilon)y^j, i) + \alpha(t + \varepsilon y^j)]\, di > \int_{\mathcal{I}} W(-t,i)\, di + \alpha t,$$

which contradicts optimality of $T^k(y;\hat{\mathbf{y}},t,\boldsymbol{\tau})$. Thus, $\tau^k = 1$ cannot be optimal, and the optimal τ^k is described by the first-order condition (A.18), which after rearranging, yields expression (4.2).

We can derive the necessary condition for optimality of τ^l with $l < k$ analogously, by differentiating (Opt$_{\text{fin}}$) with respect to τ^l. By definition of a simple tax schedule, $\tau^l < \tau^{l+1}$ and, if $l > 1$,

$\tau^l > \tau^{l-1}$. This implies that either $\tau^l = 0$ or the first-order condition has to hold with equality (if not, perturbing τ^l in either direction would be beneficial). This gives the formula in (4.3).

A.10 Proof of Corollary 2

Set $c_N(y) = y - T_N(y)$. Let $V^j_{0,N} \equiv \max_{y \in Y}\{c_N(y) - \Phi_N(y, \theta^j)\}$ be the baseline utility of type θ^j, and let $y^j_{0,N} \equiv \arg\max_{y \in Y} \{c_N(y) - \Phi_N(y, \theta^j)\}$ be his baseline income choice. Suppose by contradiction that $T'_\infty(y) < 1$ for all $y \in Y$. Then, for N sufficiently large, we have that $T'_N(y) < 1$ for all $y \in Y$ and that

$$V^n_{0,N} \geq y^{n-1}_{0,N} - T_N(y^{n-1}_{0,N}) - \Phi_N(y^{n-1}_{0,N}, \theta^n) > y^{n-1}_{0,N} - T_N(y^{n-1}_{0,N}) - \Phi_N(y^{n-1}_{0,N}, \theta^{n-1}) = V^{n-1}_{0,N}, \quad \text{(A.19)}$$

where the first inequality follows from type-θ^n's revealed preference and the second one from the fact that $\Phi_N(y^{n-1}_{0,N}, \theta^n) < \Phi_N(y^{n-1}_{0,N}, \theta^{n-1})$ for N sufficiently large.

Let $\bar{\tau}_N < 1$ be the left-derivative of $T_N(\bar{y})$. Let y^j_N be the worst-case deterministic income choice of type θ^j under $T_N(y)$ as defined by equation (A.16). By monotonicity of $V^j_{0,N}$ with respect to j, it holds that $y^1_N \leq ... \leq y^n_N$. Moreover, the fact that $y - T_N(y)$ is strictly increasing combined with (A.19) imply that, for N sufficiently large, $y^{n-1}_N < y^n_N$.

We show that $\bar{y} - y^n_{0,N} \to 0$ as $N \to \infty$. This is because the worker's revealed preference implies that

$$\Phi_N(\bar{y}, \theta^n) - \Phi_N(y^n_{0,N}, \theta^n) \geq c_N(\bar{y}) - c_N(y^n_{0,N}) \geq (1 - \bar{\tau}_N)(\bar{y} - y^n_{0,N}),$$

where the final inequality follows from concavity of $c_N(y)$. Thus, $0 \leq \bar{y} - y^n_{0,N} \leq (\Phi_N(\bar{y}, \theta^n) - \Phi_N(y^n_{0,N}, \theta^n))/(1 - \bar{\tau}_N) \to 0$ as $N \to \infty$. A very similar argument, shows that $y^n_{0,N} - y^n_N \to 0$ as $N \to \infty$.

To find a contradiction with optimality of T_N, consider the perturbation of T_N that involves slightly increasing $T'_N(y)$ to the right of $y^{n-1}_{0,N} < \tilde{y} < \bar{y}$. Formally, consider the perturbed tax schedule $T^\varepsilon_N(y) = T_N(y) + \varepsilon \max\{y - y^{n-1}_{0,N}, 0\}$. The following claim describes the welfare effect of this perturbation. The proof is a straightforward application of Lemma 6.

Claim 7.

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = p^n \left[(\alpha - \omega^n)(y^n_{0,N} - y^{n-1}_{0,N}) - \alpha \frac{y^n_{0,N} - y^{n-1}_{0,N} - \max\{y^n_N - y^{n-1}_{0,N}, 0\}}{1 - T'_N(y^n_N)}\right].$$

Applying the above results, for any arbitrarily small $\eta > 0$, there exists N sufficiently large to ensure that

$$\begin{aligned}\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} &\geq p^n \left[(\alpha - \omega^n)(y^n_{0,N} - y^{n-1}_{0,N}) - \alpha \frac{y^n_{0,N} - y^n_N}{1 - T'_N(y^n_N)}\right] \geq \\ p^n &\left[(\alpha - \omega^n)(\bar{y} - \eta - y^{n-1}_{0,N}) - \alpha \frac{\eta}{1 - T'_N(y^n_N)}\right] \geq \\ p^n &\left[(\alpha - \omega^n)(\bar{y} - \eta - \tilde{y}) - \alpha \frac{\eta}{1 - T'_N(y^n_N)}\right] > 0,\end{aligned}$$

where the strict inequality follows from Assumption 6 that implies that $\omega^n < \alpha$. Thus, $\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} > 0$, which contradicts optimality of $T_N(y)$.

A.11 Proof of Proposition 6

Fix any feasible menu of taxes $\mathcal{T} \subseteq \mathbf{T}_f$. For each $T \in \mathcal{T}$, let $\tilde{T}_T(y) = y - \text{cav}(y - T(y))$, where for any $f : Y \to \mathbb{R}$, $\text{cav}f(y)$ stands for the concavification of f. Let $c_T(y) = y - T(y)$ and $\tilde{c}_T(y) = y - \tilde{T}_T(y)$. Consider the menu of convex taxes given by $\tilde{\mathcal{T}} = \bigcup_{T \in \mathcal{T}} \{\tilde{T}_T\}$. Let

$$R(\mathcal{T}|M_i) \equiv \max_{T \in \mathcal{T}, (F,\phi) \in M_i} \{\mathbb{E}_F[T(y)], \quad \text{subject to} \quad \mathbb{E}_F[y - T(y)] - \phi = V_w(\mathcal{T}|M_i)\},$$

which is tax revenue from worker i under the menu $\mathcal{T}$ and the technology $\mathbf{M}$.

Claim 8. For every $\mathbf{M} \in \mathcal{M}_p$, there exists $\mathbf{M}' \in \mathcal{M}_p$ such that, for all $i \in \mathcal{I}$, $V_w(\tilde{\mathcal{T}}|M_i) = V_w(\mathcal{T}|M_i')$ and $R(\tilde{\mathcal{T}}|M_i) = R(\mathcal{T}|M_i')$.

Proof. We follow an approach similar to Proposition S-5 in Walton and Carroll (2022).[41] For any $T \in \mathcal{T}$, let $l_T : Y \to Y$ and $u_T : Y \to Y$ denote the endpoints of the relevant intervals of concavification of c_T. By definition, l_T and u_T satisfy $l_T(y) \leq y \leq u_T(y)$ for all $y \in Y$, $\tilde{c}_T$ is affine on $[l_T(y), u_T(y)]$, and $\tilde{c}_T$ and c_T coincide at points $l_T(y)$ and $u_T(y)$.

Fix any $\mathbf{M} \in \mathcal{M}_p$, and let $\tilde{T}_i \in \tilde{\mathcal{T}}$ and $(F_i, \phi_i) \in M_i$ be the tax and income chosen by worker i under the menu $\tilde{\mathcal{T}}$. Let $T_i \in \mathcal{T}$ be such that $\tilde{T}_{T_i} = \tilde{T}_i$. Let Y_i be the random variable with distribution F_i, and consider the income lottery F_i' and associated random variable Y_i' whose distribution conditional on Y_i is given by

$$\Pr(Y_i' = l_{T_i}(y)|Y_i = y) = \frac{u_{T_i}(y) - y}{u_{T_i}(y) - l_{T_i}(y)}, \quad \Pr(Y_i' = u_{T_i}(y)|Y_i = y) = \frac{y - l_{T_i}(y)}{u_{T_i}(y) - l_{T_i}(y)},$$

setting $\Pr(Y_i' = y|Y_i = y) = 1$ whenever $u_{T_i}(y) - l_{T_i}(y) = 0$. Observe that, for all $y \in Y$, $\mathbb{E}[Y_i'|Y_i = y] = y$ and therefore F_i' is a mean-preserving spread of F_i. Moreover, since $\tilde{c}_{T_i}$ is affine on $[l_{T_i}(y), u_{T_i}(y)]$, it holds that $\mathbb{E}[\tilde{T}_i(Y_i')|Y_i = y] = \tilde{T}_i(y)$ for all $y \in Y$, and thus $\mathbb{E}[\tilde{T}_i(Y_i')] = \mathbb{E}[\tilde{T}_i(Y_i)]$. Given this, consider the technology $\mathbf{M}' \in \mathcal{M}_p$ defined by

$$M_i' \equiv M_i \cup \{(F_i', \phi_i)\}.$$

Let us derive workers' choices when they face M_i' and the menu $\mathcal{T}$. First, we have that

$$V_w(\tilde{\mathcal{T}}|M_i) \leq V_w(\tilde{\mathcal{T}}|M_i') = \mathbb{E}[\tilde{T}_i(Y_i)] - \phi_i = V_w(\tilde{\mathcal{T}}|M_i),$$

where the two equalities follow from the fact that $\mathbb{E}[\tilde{T}_i(Y_i')] = \mathbb{E}[\tilde{T}_i(Y_i)]$ and $\mathbb{E}[Y_i'] = \mathbb{E}[Y_i]$. Moreover, since $\tilde{c}_T(y) \geq c_T(y)$ for all $T \in \mathcal{T}$ and $y \in Y$, it holds that $V_w(\mathcal{T}|M_i') \leq V_w(\tilde{\mathcal{T}}|M_i')$. Since $\tilde{c}_{T_i}$ and c_{T_i} coincide on the support of F_i', we have that

$$V_w(\mathcal{T}|M_i') \geq \mathbb{E}[c_{T_i}(Y_i')] - \phi_i = \mathbb{E}[\tilde{c}_{T_i}(Y_i)] - \phi_i = V_w(\tilde{\mathcal{T}}|M_i') \geq V_w(\mathcal{T}|M_i'),$$

and thus (F_i', ϕ_i) and T_i are optimal for the worker under M_i' and $\mathcal{T}$. This implies that $V_w(\mathcal{T}|M_i') = V_w(\tilde{\mathcal{T}}|M_i)$ as stated in the claim.

Moreover, for any choice $(F, \phi) \in M_i'$ and $T \in \mathcal{T}$ which is optimal for the worker under M_i' and $\mathcal{T}$, it must be that

$$V_w(\mathcal{T}|M_i') = \mathbb{E}_F[c_T(y)] - \phi \leq \mathbb{E}_F[\tilde{c}_T(y)] - \phi \leq V_w(\tilde{\mathcal{T}}|M_i') = V_w(\mathcal{T}|M_i'),$$

and thus $\mathbb{E}_F[c_T(y)] = \mathbb{E}_F[\tilde{c}_T(y)]$. This, together with the fact that $c_T(y) \leq \tilde{c}_T(y)$ for all $y \in Y$, in turn requires that $\tilde{c}_T(y) = c_T(y)$ for all $y \in \text{supp}(F)$, and thus, $\mathbb{E}_F[T(y)] = \mathbb{E}_F[\tilde{T}_T(y)]$ for any

[41] This approach can also be used to show convexity of the optimal singleton tax in Theorem 1. However, our original approach allows us to derive further characteristics of the optimal tax which are used later in the paper, like double monotonicity and piecewise linearity in the finite-type case.

worker-optimal choice under M_i' and $\mathcal{T}$. The fact that the worker breaks indifference in favor of the planner and that he chooses (F_i, ϕ_i) when facing M_i and $\tilde{\mathcal{T}}$, implies that (F_i', ϕ_i) and T_i is a revenue-maximizing income and tax choice among those that are optimal for the worker under M_i' and $\mathcal{T}$, and therefore

$$R(\mathcal{T}|M_i') = \mathbb{E}_{F_i'}[T_i(y)] = \mathbb{E}_{F_i'}[\tilde{T}_i(y)] = \mathbb{E}_{F_i}[\tilde{T}_i(y)] = R(\tilde{\mathcal{T}}|M_i).$$

□

Applying Claim 8,

$$V_P(\mathcal{T}) \leq V_P(\mathcal{T}|\mathbf{M}') = \int_{\mathcal{I}} [W(V_w(\mathcal{T}|M_i'), i) + \alpha R_i(\mathcal{T}|M_i')]\, di =$$

$$\int_{\mathcal{I}} [W(V_w(\tilde{\mathcal{T}}|M_i), i) + \alpha R_i(\tilde{\mathcal{T}}|M_i)]\, di = V_P(\tilde{\mathcal{T}}|\mathbf{M}).$$

Since $\mathbf{M}$ is arbitrary, this establishes that $V_P(\tilde{\mathcal{T}}) \geq V_P(\mathcal{T})$ as desired.

A.12 Proof of Proposition 7

Fix a menu of tax rules $\mathcal{T}$. For each $M \in \text{range}(\mathbf{M}^0)$, let $T_M^0 \in \mathcal{T}$ and $(F_M^0, \phi_M^0) \in M$ be an optimal choice (of tax and income) under the menu $\mathcal{T}$ and the baseline technology $\mathbf{M}^0$ for all workers i such that $M_i^0 = M$. Let $\mathcal{T}_0(\mathcal{T}) = \bigcup_{M \in \text{range}(\mathbf{M}^0)} \{T_M^0\}$. This menu of taxes is finite (by finiteness of the range of $\mathbf{M}^0$) and therefore, compact. We will show that $V_P(\mathcal{T}_0(\mathcal{T})) \geq V_P(\mathcal{T})$.

Fix a technology $\mathbf{M} \in \mathcal{M}_p$, and let $(F_i, \phi_i) \in M_i$ and $T_i \in \mathcal{T}_0(\mathcal{T})$ be worker i's choice when facing M_i and $\mathcal{T}_0(\mathcal{T})$. Put $T_i' \in \arg\min_{T \in \mathcal{T}} \mathbb{E}_{F_i}[T(y)]$. Consider the following technology $\mathbf{M}' \in \mathcal{M}_p$:

$$M_i' \equiv M_i^0 \cup \{(F_i, \phi_i + \mathbb{E}_{F_i}[T_i(y) - T_i'(y)])\}.$$

We argue that $V_P(\mathcal{T}|\mathbf{M}') \leq V_P(\mathcal{T}_0(\mathcal{T})|\mathbf{M})$.

To that end, let us characterize workers' choices under $\mathcal{T}$ and $\mathbf{M}'$. Observe that, if worker i chooses $(F_i, \phi_i + \mathbb{E}_{F_i}[T_i(y) - T_i'(y)]) \in M_i'$, the optimal associated tax choice in $\mathcal{T}$ is T_i'. Thus, worker i's payoff from making this income choice is

$$\mathbb{E}_{F_i}[y - T_i'(y)] - \phi_i - \mathbb{E}_{F_i}[T_i(y) - T_i'(y)] = \mathbb{E}_{F_i}[y - T_i(y)] - \phi_i = V_w(\mathcal{T}_0(\mathcal{T})|M_i).$$

Suppose first that $V_w(\mathcal{T}_0(\mathcal{T})|M_i) = V_w(\mathcal{T}|M_i^0)$. Then, both when facing $\mathcal{T}$ and M_i', and when facing $\mathcal{T}_0(\mathcal{T})$ and M_i, worker i's indirect utility is the same. When facing $\mathcal{T}$ and M_i', worker i can achieve his optimal utility by either picking the best income choice in M_i^0 and the tax $T_{M_i^0}^0$, or by picking $(F_i, \phi_i + \mathbb{E}_{F_i}[T_i(y) - T_i'(y)])$ and the tax T_i'. In both cases this choice leads to lower tax revenue than what obtains under $\mathcal{T}_0(\mathcal{T})$ and M_i: In the former case, this holds because of the worker's favorable tie-breaking, and in the latter case because of the definition of T_i' which implies that $\mathbb{E}_{F_i}[T_i'(y)] \leq \mathbb{E}_{F_i}[T_i(y)]$. Therefore, total welfare conditional on the event $V_w(\mathcal{T}_0(\mathcal{T})|M_i) = V_w(\mathcal{T}|M_i^0)$ is weakly lower under $\mathcal{T}$ and M_i' than under $\mathcal{T}_0(\mathcal{T})$ and M_i.

Second, suppose that $V_w(\mathcal{T}_0(\mathcal{T})|M_i) > V_w(\mathcal{T}|M_i^0) = V_w(\mathcal{T}_0(\mathcal{T})|M_i^0)$. Then, worker i's uniquely optimal choice when facing M_i' and $\mathcal{T}$ is $(F_i, \phi_i + \mathbb{E}_{F_i}[T_i(y) - T_i'(y)])$ and the tax T_i'. As argued above, this gives him a payoff of $\mathbb{E}_{F_i}[y - T_i(y)] - \phi_i = V_w(\mathcal{T}_0(\mathcal{T})|M_i)$, which by assumption is strictly greater than his optimal payoff when choosing an element of M_i^0. Moreover, expected tax revenue from the worker's optimal choice under $\mathcal{T}$ and M_i' is $\mathbb{E}_{F_i}[T_i'(y)] \leq \mathbb{E}_{F_i}[T_i(y)]$, where the inequality follows again from the definition of T_i'.

Combining everything and summing over workers, we have

$$V_P(\mathcal{T}) \leq V_P(\mathcal{T}|\mathbf{M}') \leq \int_{\mathcal{I}} [W(V_w(\mathcal{T}_0(\mathcal{T})|M_i), i) + \alpha \mathbb{E}_{F_i}[T_i(y)]]\, di = V_P(\mathcal{T}_0(\mathcal{T})|\mathbf{M}).$$

Since $\mathbf{M}$ was arbitrary, it follows that $V_P(\mathcal{T}) \leq V_P(\mathcal{T}_0(\mathcal{T}))$.

A.13 Proof of Proposition 8

The proof, which is a straightforward generalization of Theorem 1, is in Online Appendix OA.2.7.

B Differentiability of worst-case welfare with respect to tax perturbations

In this section, we establish that worst-case welfare is directionally differentiable at certain tax rules and provide an expression for the derivative. The result can be used to derive necessary conditions for worst-case optimality and is later applied in some of our proofs in Appendix A. The proof is provided in Online Appendix OA.2.8.

Take any $T \in \mathbf{T}_f$ that is convex. By convexity, T is directionally differentiable. We let $T'(y+)$ and $T'(y-)$ denote respectively its right- and left-derivatives. Consider the perturbation

$$T^\varepsilon(y) = T(y) + \varepsilon D(y),$$

where $D : Y \to \mathbb{R}$ is continuous and $\varepsilon \in \mathbb{R}$. Let $y_i \equiv \min\{y \in Y : V_w(T|M_i^0) \leq y - T(y)\}$, which is well-defined given that $T(y)$ is continuous and $V_w(T|M_i^0) \leq \overline{y} - T(\overline{y})$. Let $\hat{y} \in Y$ be the highest y such that $T'(y-) = 0$ and set $\hat{y} = 0$ if $T'(0+) > 0$. Let $R_i(\varepsilon) \equiv \mathbb{E}_{F_i^\varepsilon}[T^\varepsilon(y)]$, where $F_i^\varepsilon \in \Delta(Y)$ is worker i's worst-case income choice under T^ε as defined in Lemma 1.

Lemma 6. *Let $T \in \mathbf{T}_f$ be any convex, double-monotone, strictly increasing tax rule. Then, $\varepsilon \to V_P(T^\varepsilon)$ is directionally differentiable at $\varepsilon = 0$, with derivatives*[42] *given by*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0} = \int_{\mathcal{I}} \left[W_1(V_w(T|M_i^0), i) \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon *}\right|_{\varepsilon=0} + \alpha \left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0} \right] di, \quad * \in \{+,-\}, \tag{B.1}$$

where

$$\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon +} = \max_{(F,\phi) \in X_w(T^\varepsilon|M_i^0)} \mathbb{E}_F[-D(y)], \quad \frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon -} = \min_{(F,\phi) \in X_w(T^\varepsilon|M_i^0)} \mathbb{E}_F[-D(y)], \tag{B.2}$$

$$\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = - \min_{\lambda \in \Lambda_i^*(0)} \max_{F \in X_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon +}\right|_{\varepsilon=0} \right\} \tag{B.3}$$

$$\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon -}\right|_{\varepsilon=0} = - \max_{\lambda \in \Lambda_i^*(0)} \min_{F \in X_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon -}\right|_{\varepsilon=0} \right\}, \tag{B.4}$$

and

$$\Lambda_i^*(0) = \left[\frac{T'(y_i-)}{1 - T'(y_i-)}, \frac{T'(y_i+)}{1 - T'(y_i+)} \right],$$

$$X_i^*(0) = \begin{cases} \{F \in \Delta(Y) : \mathbb{E}_F[T(y)] = T(y_i), \mathbb{E}_F[y] = y_i\}, & \textit{if } y_i > \hat{y}, \\ \{F \in \Delta([0,\hat{y}]) : \mathbb{E}_F[y] \geq y_i\}, & \textit{if } y_i \leq \hat{y}. \end{cases}$$

[42] If $\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +} = \frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon -}$, then they are both equal to the Gateaux differential of V_P at T in the direction D.

Online Appendix

OA.1 Simple tax schedules

For any $m \leq n$, any vector of income cutoffs $\hat{\mathbf{y}} \in Y^{m-1}$, and marginal tax rates $\boldsymbol{\tau} \in [0,1]^m$, let $\boldsymbol{\tau}_{\text{nr}}(\hat{\mathbf{y}}, \boldsymbol{\tau}) \in [0,1]^k$ with $k \leq m$ be the associated vector of *non-redundant marginal tax rates*, satisfying

$$\bigcup_{l=1}^{k}\{\tau^l_{\text{nr}}(\hat{\mathbf{y}},\boldsymbol{\tau})\} = \bigcup_{l=1}^{m}\{\tau^l : y^j \in [\hat{y}^{l-1},\hat{y}^l] \text{ for some } j \in \{1,...,n\}\}$$

and $\tau^1_{\text{nr}}(\hat{\mathbf{y}},\boldsymbol{\tau}) < ... < \tau^k_{\text{nr}}(\hat{\mathbf{y}},\boldsymbol{\tau})$.

Definition 3 (Simple tax schedule)**.** The tax schedule T is simple if there exist $(k, \hat{\mathbf{y}}, t, \boldsymbol{\tau}) \in \{1,...,n\} \times Y^{k-1} \times \mathbb{R}_- \times [0,1]^k$ with $\hat{y}^1 < ... < \hat{y}^{k-1}$ such that $T(y) = T^k(y; \hat{\mathbf{y}}, t, \boldsymbol{\tau})$ as given by (4.1) and

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{\text{nr}}(\hat{\mathbf{y}},\boldsymbol{\tau}).$$

OA.2 Omitted proofs

OA.2.1 Proof of Claim 1

First, we argue that the minimum in (A.2)-(A.3) is attained. Let $C_i \equiv \max_{F \in \Delta(Y)}\{\mathbb{E}_F[y - T(y)] - V_w(T|M_i^0)\}$, which is well-defined given compactness of $\Delta(Y)$. Then, without loss we can restrict the domain of minimization to the compact set $(F, \phi) \in \Delta(Y) \times [0, C_i]$. Because the objective is continuous in (F, ϕ), the minimum is attained.

Second, we show that without loss of optimality, (A.3) holds with equality. Let $X_i^* \subseteq \Delta(Y) \times \mathbb{R}_+$ be the set of income choices that attains the minimum for worker i in (A.2)-(A.3). Let $(F_i, \phi_i) \in X_i^*$, and consider $(F_i, \phi_i') \in \Delta(Y) \times \mathbb{R}_+$ such that $\phi_i' = \mathbb{E}_{F_i}[y - T(y)] - V_w(T|M_i^0)$. (F_i, ϕ_i') satisfies (A.3) with equality and, by non-decreasingness of $W(\cdot, i)$, attains a weakly lower value than (F_i, ϕ_i). Thus, $(F_i, \phi_i') \in X_i^*$.

Finally, we show the equality stated in the claim. Let $\tilde{v}_i(T)$ be the value of (A.2)-(A.3). For any $(F_i, \phi_i) \in X_i^*$, it holds that $\mathbb{E}_{F_i}[y - T(y)] \geq V_w(T|M_i^0)$, and thus $r_i(T) \leq \mathbb{E}_{F_i}[T(y)]$. On the other hand, letting F_i be an argmin for $r_i(T)$, we can define $(F_i, \mathbb{E}_{F_i}[y - T(y)] - V_w(T|M_i^0)) \in \Delta(Y) \times \mathbb{R}_+$, which is feasible in (A.2)-(A.3). Hence, $r_i(T) \geq \mathbb{E}_{F_i'}[T(y)]$ for F_i' such that $(F_i', \phi_i) \in X_i^*$. This establishes the claim.

OA.2.2 Proof of Lemma 3

Take any feasible T such that $c(y) = y - T(y)$ is non-decreasing and suppose that $T(0) < -\overline{C}$ (otherwise, the result holds trivially) and consider a feasible perturbation of T given by $T^\varepsilon(y) = T(y) + \varepsilon$ with $\varepsilon > 0$. Since a constant shift in $T(y)$ does not affect workers' worst-case income choice as defined in Lemma 1, the function $\varepsilon \to V_P(T^\varepsilon)$ is differentiable and we can write

$$\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon} = -\int_{\mathcal{I}} W_1(V_w(T|M_i^0) - \varepsilon, i)\, di + \alpha. \tag{OA.1}$$

By monotonicity of $c(y)$, for all $i \in \mathcal{I}$, it holds that $V_w(T|M_i^0) \geq c(0) - \overline{\phi} > \overline{C} - \overline{\phi} = C$, where the second inequality follows from the assumption that $c(0) > \overline{C}$. Then, Assumption 2 implies that $\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon} \geq 0$ whenever $c(0) > \overline{C}$. Thus, T^ε with $\varepsilon > 0$ is a weak improvement over T. In particular, the tax rule $\hat{T}(y) = T(y) + c(0) - \overline{C}$ is feasible and satisfies $V_P(\hat{T}) \geq V_P(T)$.

OA.2.3 Proof of Claim 3

Let $T^k \in \hat{\mathbf{T}}$ be a sequence of tax rules converging to $T^\infty \in \hat{\mathbf{T}}$. Fix a plausible technology $\mathbf{M} \in \mathcal{M}_p$ and let $(F_i^k, \phi_i^k) \in M_i$ be worker i's chosen action under T^k and M_i. Without loss, $V_P(T^k)$ converges (if not, replace T^k by a subsequence). By compactness of M_i, without loss, for every $i \in \mathcal{I}$, (F_i^k, ϕ_i^k) converges to an element of M_i, which we denote by $(F_i^\infty, \phi_i^\infty)$. By continuity, $(F_i^\infty, \phi_i^\infty)$ is optimal for worker i under T^∞ and M_i. Since the worker breaks indifference in favor of the planner, the planner's payoff under T^∞ and $\mathbf{M}$ is bounded below by what obtains if all workers choose $(F_i^\infty, \phi_i^\infty)$. Thus,[43]

$$\begin{aligned} V_P(T^\infty|\mathbf{M}) &\geq \int_{\mathcal{I}} \{W(\mathbb{E}_{F_i^\infty}[y - T^\infty(y)] - \phi_i^\infty, i) + \alpha \mathbb{E}_{F_i^\infty}[T^\infty(y)]\}\, di && \text{(OA.2)} \\ &= \lim_{k\to\infty} V_P(T^k|\mathbf{M}) \geq \lim_{k\to\infty} V_P(T^k). && \text{(OA.3)} \end{aligned}$$

Since $\mathbf{M}$ was arbitrary, $V_P(T^\infty) \geq \lim_{k\to\infty} V_P(T^k)$, which establishes the result.

OA.2.4 Proof of Claim 5

The fact that $\tau < 1$ implies that $y - T(y)$ is strictly increasing and thus Lemma 6 applies. We begin by computing the directional derivative for the case with $V_w(T|M_i^0) < \overline{y} - T(\overline{y})$. For any such i, $\Lambda_i^*(0) = \{\tau/(1-\tau)\}$. Fix $(F_i^0, \phi_i^0) \in M_i^0$ to be an optimal income choice for type i under M_i^0 and the tax T, and put $y_i^0 = \mathbb{E}_{F_i^0}[y]$. Then, under an affine tax, $y_i = y_i^0 - \phi_i^0/(1-\tau)$. If $\tau = 0$, then $\hat{y} = \overline{y}$, and thus

$$X_i^*(0) = \{F \in \Delta(Y) : \mathbb{E}_F[y] \geq y_i\}.$$

And hence by Lemma 6,

$$\begin{aligned} \left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} &= -\max_{F\in\Delta(Y)} \{\mathbb{E}_F[\min\{\tilde{y} - y, 0\}], \quad \text{subject to} \quad \mathbb{E}_F[y] \geq y_i\} = \\ & -\max_{y\in Y}\{\min\{\tilde{y} - y, 0\}, \quad \text{subject to} \quad y \geq y_i\} = -\min\{\tilde{y} - y_i, 0\}, \end{aligned}$$

where the second equality follows from concavity of $y \to \min\{\tilde{y} - y, 0\}$.

Second if $\tau \in (0, 1)$, then $\hat{y} = 0$, and thus

$$X_i^*(0) = \{F \in \Delta(Y) : \mathbb{E}_F[y] = y_i\}.$$

Applying Lemma 6,

$$\begin{aligned} \left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} &= -\frac{1}{1-\tau} \max_{F\in\Delta(Y):\mathbb{E}_F[y]=y_i} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] + \frac{\tau}{1-\tau} \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}\right|_{\varepsilon=0} = \\ & -\frac{1}{1-\tau} \min\{\tilde{y} - y_i, 0\} + \frac{\tau}{1-\tau} \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}\right|_{\varepsilon=0}, \end{aligned}$$

where the second equality follows again from concavity of $y \to \min\{\tilde{y} - y, 0\}$.

[43] The equality in (OA.2) uses the Dominated Convergence Theorem, which can be applied due to the fact that for each $k \in \mathbb{N}$ and $i \in \mathcal{I}$:

$$|W(\mathbb{E}_{F_i^k}[y - T^k(y)] - \phi_i^k, i) + \alpha \mathbb{E}_{F_i^k}[T^k(y)]| \leq |W(\overline{C} + \overline{y}), i)| + \alpha \overline{y},$$

which is an integrable function of i.

If $V_w(T|M_i^0) = \bar{y} - T(\bar{y})$, then $y_i = \bar{y}$ and Lemma 6 implies that

$$\frac{\partial R_i(\varepsilon)}{\partial \varepsilon+}\bigg|_{\varepsilon=0} = -\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}\bigg|_{\varepsilon=0} = -\min\{\tilde{y} - \bar{y}, 0\}.$$

Applying (B.1) and the definition of $\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}$, we then obtain the expression in Lemma 5.

OA.2.5 Proof of Claim 6

For an affine T with $\tau \in [0,1)$, the resulting consumption rule $y - T(y)$ is strictly increasing. As a result, $V_w(T|M_i^0) = \max_{y \in Y}\{y - T(y)\} \iff (\delta_{\overline{y}}, 0) \in M_i^0$. For any such i, $(\delta_{\overline{y}}, 0)$ is the unique income choice $(F, \phi) \in \Delta(Y) \times \mathbb{R}_+$ satisfying $\mathbb{E}_F[y - T(y)] \geq V_w(T|M_i^0)$. Therefore, the distinction between the two cases in Lemma 1 is vacuous, and we have, applying the lemma,

$$V_P(T) = \int_{\mathcal{I}} [W(V_w(T|M_i^0), i) + \alpha r_i(T)]\, di,$$

where $r_i(T)$ is defined by (A.1).

Next, we show that $r_i(T) = \frac{\tau V_w(T|M_i^0)+t}{1-\tau}$ for all $\tau \in [0,1)$ and $i \in \mathcal{I}$. The result is immediate if $\tau = 0$, since then revenue is constant and equal to t. Suppose now that $\tau \in (0,1)$. We argue that, for any $F_i \in \Delta(Y)$ that is a solution for $r_i(T)$, it must be that $\mathbb{E}_{F_i}[y - T(y)] = V_w(T|M_i^0)$. Suppose towards a contradiction that F_i is a solution for $r_i(T)$ and that $\mathbb{E}_{F_i}[y - T(y)] > V_w(T|M_i^0)$. By Assumption 3, $V_w(T|M_i^0) \geq -T(0)$. Since by hypothesis $\mathbb{E}_{F_i}[y - T(y)] > V_w(T|M_i^0) \geq -T(0)$, it cannot be that F_i assigns probability one to $y = 0$. Then, for $\eta \in (0,1)$, consider $\tilde{F}_i = (1-\eta)F_i + \eta\delta_0$. If η is sufficiently small, $\mathbb{E}_{\tilde{F}_i}[y - T(y)] \geq V_w(T|M_i^0)$. Moreover, since $T(y)$ is strictly increasing and F_i is not degenerate at $y = 0$, we have that $\mathbb{E}_{\tilde{F}_i}[T(y)] < \mathbb{E}_{F_i}[T(y)]$. This contradicts F_i being revenue-minimizing. Thus, $\mathbb{E}_{F_i}[y - T(y)] = V_w(T|M_i^0)$ and we can write

$$r_i(T) = \mathbb{E}_{F_i}[T(y)] = \frac{\tau \mathbb{E}_{F_i}[y - T(y)] + t}{1-\tau} = \frac{\tau V_w(T|M_i^0) + t}{1-\tau}, \quad \forall i \in \mathcal{I}.$$

Let us now consider the case in which $\tau = 1$. If $\phi_i^0(\tau) > 0$, then worst-case revenue is given by $r_i(T)$, which is continuous in τ and by the previous step converges to $-t$ as $\tau \to 1$. If $\phi_i^0(\tau) = 0$, then Lemma 1 implies that worker i's worst-case income choice under τ is given by $F_i^0(\tau)$, and thus worst-case revenue from this worker equals $t + y_i^0(\tau)$.

OA.2.6 Proof of Lemma 5

First, by Lemma 2, workers' baseline income choice is deterministic and described by its mean $y_0^j \in Y$. Second, because worst-case welfare is weakly increasing in $V_w(T|M^0(\theta))$ (see Lemma 1), we can put $\mathbf{y}^0$ as an optimization variable in the planner's problem. Third, once we restrict attention to double-monotone and convex tax schedules, it follows from Lemma 1 that the deterministic income choice defined in (A.16) attains the worst-case revenue. These three features combined yield the program in the lemma.

OA.2.7 Proof of Proposition 8

We begin by stating a version of Lemma 1 that allows for workers' risk aversion. The proof is identical to Lemma 1 so we omit it.

Lemma OA.1. *For any feasible T,*

$$V_P(T) = \int_{\mathcal{I}} [W(V_w(T|M_i^0), i) + \alpha \mathbb{E}_{F_i}[T(y)]]\, di, \tag{OA.4}$$

where F_i is defined by:

(i) If $V_w(T|M_i^0) < \max_{y\in Y} \tilde{u}(y - T(y))$*:*

$$F_i \in \arg\min_{F\in\Delta(Y)} \; \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[\tilde{u}(y - T(y))] \geq V_w(T|M_i^0),$$

(ii) If $V_w(T|M_i^0) = \max_{y\in Y} \tilde{u}(y - T(y))$*:*

$$F_i \in \arg\max_{(F,0)\in M_i^0} \; \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[\tilde{u}(y - T(y))] = V_w(T|M_i^0).$$

The rest of the proof consists of modifying the arguments in the proof of Theorem 1. To avoid repetition, here we only focus on the arguments that are not exactly the same as in the original proof. Let T be any feasible tax rule, and let $u(y) = \tilde{u}(y - T(y))$ be the associated utility contract. To construct the relevant separating hyperplanes, consider the sets

$$A = \text{co}\{(u(y), y - \tilde{u}^{-1}(u(y))) : y \in Y\}, \quad B_i = \{(u, v) : u > V_w(T|M_i^0), v < \mathbb{E}_{F_i}[T(y)]\},$$

where F_i is defined as in Lemma OA.1. For every $i \in \mathcal{I}$, A and B_i are convex and disjoint, and thus the Separating Hyperplane Theorem implies that we can construct a i-specific utility contract, u_i, implicitly defined by:

$$\kappa_i + \lambda_i u_i(y) - \beta_i(y - \tilde{u}^{-1}(u_i(y))) = 0, \tag{OA.5}$$

where $(\kappa_i, \beta_i, \lambda_i) \in \mathbb{R} \times \mathbb{R}_+^2$ satisfy for all $i \in \mathcal{I}$

$$\kappa_i + \lambda_i u(y) - \beta_i(y - \tilde{u}^{-1}(u(y))) \leq 0, \quad \forall y \in Y, \tag{OA.6}$$

$$\kappa_i + \lambda_i \mathbb{E}_{F_i}[u(y)] - \beta_i \mathbb{E}_{F_i}[y - \tilde{u}^{-1}(u(y))] = 0. \tag{OA.7}$$

Claim OA.1. *If $\tilde{u}$ is concave on $\mathbb{R}_+$, then for all $i \in \mathcal{I}$, u_i is continuous, non-decreasing and concave, and satisfies for all $y \in Y$, $u_i(y) \geq u(y)$.*

Proof. The Claim is immediate if $\beta_i = 0$, so suppose that $\beta_i > 0$. First, we begin by showing that (OA.5) has a unique solution $u_i(y)$ for each $y \in Y$ and $i \in \mathcal{I}$. By (OA.6), for any $y \in Y$, there exist $u \geq \overline{u}$ (e.g., $u = u(y)$) such that $\kappa_i + \lambda_i u - \beta_i(y - \tilde{u}^{-1}(u)) \leq 0$. On the other hand, setting $u = \tilde{u}(c)$ where $c \in \mathbb{R}_+$ is arbitrarily large, we have that

$$\kappa_i + \lambda_i u - \beta_i(y - \tilde{u}^{-1}(u)) = \kappa_i + \lambda_i \tilde{u}(c) - \beta_i(y - c) \geq 0,$$

where the inequality follows from the fact that $\beta_i > 0$. Hence, by continuity, there exists $u \geq \underline{u}$ that satisfies (OA.5) with equality. Uniqueness of the solution follows from the fact that the left-hand-side of (OA.5) is strictly increasing in $u_i(y)$.

Second, to show that $u_i(y)$ is non-decreasing, take $y, y' \in Y$ such that $y' > y$. (OA.5) implies that

$$\lambda_i(u_i(y') - u_i(y)) + \beta_i(\tilde{u}^{-1}(u_i(y')) - \tilde{u}^{-1}(u_i(y))) = \beta_i(y' - y) > 0, \tag{OA.8}$$

where the inequality follows from the fact that $\beta_i > 0$. Given that $(\lambda_i, \beta_i) \neq (0, 0)$, $\lambda_i \geq 0$ and $\tilde{u}^{-1}$ is strictly increasing, (OA.8) requires that $u_i(y') \geq u_i(y)$.

Third, we show that $u_i(y)$ is concave. Suppose first that $\lambda_i > 0$ and suppose toward a contradiction that there is $y, y' \in Y$ with $y \neq y'$ and $\gamma \in (0, 1)$ such that $\gamma u_i(y) + (1 - \gamma)u_i(y') > u_i(y_\gamma)$, where $y_\gamma = \gamma y + (1 - \gamma)y'$. By the definition of $u_i(y)$, we have

$$u_i(y_\gamma) - \gamma u_i(y) - (1 - \gamma)u_i(y') = \frac{\beta_i}{\lambda_i}[\gamma\tilde{u}^{-1}(u_i(y)) + (1 - \gamma)\tilde{u}^{-1}(u_i(y')) - \tilde{u}^{-1}(u_i(y_\gamma))] >$$
$$\frac{\beta_i}{\lambda_i}[\gamma\tilde{u}^{-1}(u_i(y)) + (1 - \gamma)\tilde{u}^{-1}(u_i(y')) - \tilde{u}^{-1}(\gamma u_i(y) + (1 - \gamma)u_i(y'))] \geq 0,$$

where the first inequality follows from the assumption that $\gamma u_i(y) + (1-\gamma)u_i(y') > u_i(y_\gamma)$ and the fact that $\tilde{u}^{-1}$ is strictly increasing, and the second inequality from convexity of $\tilde{u}^{-1}$. This contradicts $\gamma u_i(y) + (1-\gamma)u_i(y') > u_i(y_\gamma)$. Second, if $\lambda_i = 0$, then

$$u_i(y) = \tilde{u}(-\kappa_i/\beta_i + y),$$

which is concave by concavity of $\tilde{u}(y)$.

Fourth, we show that $u_i(y)$ is continuous. The fact that $\tilde{u}$ is strictly increasing implies that it is almost everywhere differentiable. Let $C \subset (0, +\infty)$ be an open set in which $\tilde{u}$ is differentiable. (OA.5) is strictly increasing in $u_i(y)$, and therefore by the Implicit Function Theorem, u_i is differentiable (and therefore continuous) on C. By concavity of u_i this is enough to ensure that u_i is continuous on Y (see Theorem 5.27 in Aliprantis and Border (2006)).

Finally, for each $y \in Y$, (OA.5) and (OA.6) imply that

$$\lambda_i(u_i(y) - u(y)) + \beta_i[\tilde{u}^{-1}(u_i(y)) - \tilde{u}^{-1}(u(y))] \geq 0,$$

which is only possible if $u_i(y) \geq u(y)$. □

Let $\overline{u}(y) = \inf_{i \in \mathcal{I}} u_i(y)$, and $\overline{T}(y) = y - \tilde{u}^{-1}(\overline{u}(y))$. By Claim OA.1, $\overline{u}$ is feasible, non-decreasing, concave, and satisfies $V_w(\overline{T}|M_i^0) \geq V_w(T|M_i^0)$ for all $i \in \mathcal{I}$. An analogous argument to the one in the proof of Theorem 1, shows that $\overline{T}$ yields weakly higher expected revenue than T conditional on every $i \in \mathcal{I}$. Thus, $V_P(\overline{T}) \geq V_P(T)$. Existence of the maximum can also be established analogously to Theorem 1.

OA.2.8 Proof of Lemma 6

Step 1: Envelope theorem for problems with concave parametrized constraints. We begin by providing an extension of the envelope theorem for problems with paramatrized constraints by Milgrom and Segal (2002) (henceforth, MS), that does not require the constraint to be differentiable—as assumed in their Theorem 5 and Corollary 5—, but instead requires that the constraint is concave with respect to the parameter.

As MS, we are concerned with a constrained optimization problem of the form

$$V(t) = \sup_{x \in X : g(x,t) \geq 0} f(x,t), \quad X^*(t) = \{x \in X : g(x,t) \geq 0, f(x,t) = V(t)\}, \tag{OA.9}$$

where X is a convex compact set in a normed linear space, and $f : X \times [\underline{t}, \overline{t}] \to \mathbb{R}$ and $g : X \times [\underline{t}, \overline{t}] \to \mathbb{R}^k$ and $\underline{t}, \overline{t} \in \mathbb{R}$ satisfy $\underline{t} < \overline{t}$. We define the Lagrangian of this problem as $L : X \times \mathbb{R}_+^k \times [\underline{t}, \overline{t}] \to \mathbb{R}$

$$L(x,y,t) = f(x,t) + \sum_{i=1}^{k} y_i g_i(x,t),$$

and we let $Y^*(t)$ be the set of solutions to the dual program,

$$Y^*(t) = \arg\min_{y \in \mathbb{R}_+^k} \sup_{x \in X} L(x,y,t).$$

Lemma OA.2. *Suppose that f and g are continuous and concave in x, $f_t(x,t)$ is continuous in (x,t), $g(x,t)$ is continuous and concave in t for all $x \in X$ with directional derivatives with respect to t that are continuous in x, and there exists $\hat{x} \in X$ such that $g(\hat{x},t) \gg 0$ for all $t \in [\underline{t}, \overline{t}]$. Then, V is directionally differentiable and its directional derivatives equal:*

$$V'(t+) = \max_{x \in X^*(t)} \min_{y \in Y^*(t)} \frac{\partial L(x,y,t)}{\partial t+} = \min_{y \in Y^*(t)} \max_{x \in X^*(t)} \frac{\partial L(x,y,t)}{\partial t+}, \quad \textit{for } t < \overline{t}$$

$$V'(t-) = \min_{x \in X^*(t)} \max_{y \in Y^*(t)} \frac{\partial L(x,y,t)}{\partial t-} = \max_{y \in Y^*(t)} \min_{x \in X^*(t)} \frac{\partial L(x,y,t)}{\partial t-}, \quad \textit{for } t > \underline{t}.$$

Proof. By an analogous argument to Corollary 5 in MS, it holds that $X^*(t) \times Y^*(t)$ is non-empty and equal to the saddle-set of the Lagrangian, and thus

$$V(t) = \max_{x \in X} \min_{y \in \mathbb{R}^k_+} L(x, y, t) = \min_{y \in \mathbb{R}^k_+} \max_{x \in X} L(x, y, t).$$

Fix $t_0 \in [\underline{t}, \bar{t})$, by definition of the saddle point, for any selection $(x(t), y(t)) \in X^*(t) \times Y^*(t)$ and any $t > t_0$, we can write

$$\frac{L(x(t_0), y(t), t) - L(x(t_0), y(t), t_0)}{t - t_0} \leq \frac{V(t) - V(t_0)}{t - t_0} \leq \frac{L(x(t), y(t_0), t) - L(x(t), y(t_0), t_0)}{t - t_0}. \tag{OA.10}$$

The fact that $g_i(x, \cdot)$ is concave implies that it is directionally differentiable and thus, by the generalized Mean Value Theorem,

$$\frac{\partial g_i(x(t_0), t'(t))}{\partial t+} \leq \frac{g_i(x(t_0), t) - g(x(t_0), t_0)}{t - t_0} \leq \frac{\partial g_i(x(t_0), t'(t))}{\partial t-}.$$

for some $t'(t) \in [t_0, t]$. Moreover, concavity of $g_i(x, \cdot)$ implies that

$$\frac{g_i(x(t), t) - g_i(x(t), t_0)}{t - t_0} \leq \frac{\partial g_i(x(t), t_0)}{\partial t_0+}.$$

Further, differentiability of $f(x, \cdot)$ and the Mean Value Theorem imply that

$$\frac{f(x(t_0), t) - f(x(t_0), t_0)}{t - t_0} = f_t(x(t_0), t''(t)), \quad \frac{f(x(t), t) - f(x(t), t_0)}{t - t_0} = f_t(x(t), t'''(t)),$$

for some $t''(t), t'''(t) \in [t_0, t]$.

Combining all of this, we have

$$\max_{x \in X^*(t_0)} \left\{ f_t(x, t''(t)) + \sum_{i=1}^k y_i(t) \frac{\partial g_i(x, t'(t))}{\partial t+} \right\} \leq \frac{V(t) - V(t_0)}{t - t_0} \leq \min_{y \in Y^*(t_0)} \left\{ f_t(x(t), t'''(t)) + \sum_{i=1}^k y_i \frac{\partial g_i(x(t), t_0)}{\partial t+} \right\}.$$

And, in the limit,

$$\liminf_{t \downarrow t_0} \max_{x \in X^*(t_0)} \left\{ f_t(x, t''(t)) + \sum_{i=1}^k y_i(t) \frac{\partial g_i(x, t'(t))}{\partial t+} \right\} \geq \liminf_{t \downarrow t_0} \max_{x \in X^*(t_0)} \left\{ f_t(x, t''(t)) + \sum_{i=1}^k y_i(t) \frac{\partial g_i(x, t_0)}{\partial t+} \right\} \geq \min_{y \in Y^*(t_0)} \max_{x \in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^k y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\},$$

where the first inequality follows from lower semi-continuity of the directional derivatives of $g(x, \cdot)$ with respect to t which in turn follows from concavity of $g(x, \cdot)$ (see Theorem 33 in Fenchel and Blackett (1953)), and the last inequality follows from the fact that the Lagrangian is continuous, and hence the saddle set is upper hemicontinuous in t by Berge's Maximum Theorem.[44]

[44] To ensure compactness of the choice set in OA.12, we can bound above the relevant choices for y in the same way as in MS.

And similarly,

$$\limsup_{t\downarrow t_0} \min_{y\in Y^*(t_0)} \left\{ f_t(x(t), t'''(t)) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x(t), t_0)}{\partial t+} \right\} \leq$$
$$\max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\},$$

where the inequality follows again from upper hemicontinuity of the saddle set.

Given the above, taking the limits inferior and superior in (OA.10), we have

$$\min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\} \leq \liminf_{t\downarrow t_0} \frac{V(t) - V(t_0)}{t - t_0}$$
$$\leq \limsup_{t\downarrow t_0} \frac{V(t) - V(t_0)}{t - t_0} \leq \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\}$$

We also know that

$$\min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\} \geq \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\}.$$

Hence,

$$V'(t_0+) = \min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\} = \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t+} \right\}.$$

The argument for $V'(t-)$ is analogous, which gives

$$V'(t-) = \max_{y\in Y^*(t_0)} \min_{x\in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t-} \right\} = \min_{x\in X^*(t_0)} \max_{y\in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t-} \right\}.$$

□

Step 2: Directional derivatives of $\varepsilon \to V_w(T^\varepsilon | M_i^0)$.

Claim OA.2. *For all $i \in \mathcal{I}$, the function $\varepsilon \to V_w(T^\varepsilon | M_i^0)$ is convex.*

Proof. Take $\varepsilon, \varepsilon' \in \mathbb{R}$ and $\gamma \in [0, 1]$, and let $\varepsilon_\gamma = \gamma\varepsilon + (1-\gamma)\varepsilon'$. Applying the definition,

$$\gamma V_w(T^\varepsilon | M_i^0) + (1-\gamma) V_w(T^{\varepsilon'} | M_i^0) =$$
$$\gamma \max_{(F,\phi)\in M_i^0} \{\mathbb{E}_F[y - T^\varepsilon(y)] - \phi\} + (1-\gamma) \max_{(F,\phi)\in M_i^0} \{\mathbb{E}_F[y - T^{\varepsilon'}(y)] - \phi\} \geq$$
$$\max_{(F,\phi)\in M_i^0} \{\mathbb{E}_F[y - \gamma T^\varepsilon(y) - (1-\gamma) T^{\varepsilon'}(y)] - \phi\} = V_w(T^{\varepsilon_\gamma} | M_i^0).$$

□

It follows from Claim OA.2 that $V_w(T^\varepsilon | M_i^0)$ is directionally differentiable with respect to ε. Its directional derivative can be computed applying Corollary 4 in MS. We state its expression in the following claim.

Claim OA.3. *For all $i \in \mathcal{I}$, the function $\varepsilon \to V_w(T^\varepsilon|M_i^0)$ is directionally differentiable with derivatives*

$$\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+} = \max_{(F,\phi)\in X_w(T^\varepsilon|M_i^0)} \mathbb{E}_F[-D(y)],$$
$$\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon-} = \min_{(F,\phi)\in X_w(T^\varepsilon|M_i^0)} \mathbb{E}_F[-D(y)].$$

Step 3: Directional derivatives of $\varepsilon \to R_i(T^\varepsilon)$. Next, we turn to the revenue component of (3.1). Under conditions (i) or (ii) in Lemma 6 we have that, for all $i \in \mathcal{I}$, worst-case revenue is determined by Case (i) in Lemma 1. In particular, if there is a positive measure of types such that $V_w(T|M_i^0) = \max\limits_{y\in Y}\{y - T(y)\}$ and thus $\phi = 0$ for all $(F,\phi) \in M_i^*(0)$, strict monotonicity of T implies that there exists a unique $(F, 0)$ that attains $V_w(T|M_i^0)$, and thus the favorable tie-breaking rule does not play a role. Therefore, for ε small enough, worst-case revenue under T^ε for any i can be written as

$$R_i(\varepsilon) = \min_{F\in\Delta(Y)} \mathbb{E}_F[T^\varepsilon(y)], \quad s.t. \quad \mathbb{E}_F[y - T^\varepsilon(y)] \geq V_w(T^\varepsilon|M_i^0). \tag{OA.11}$$

We compute the directional derivatives of $R_i(\varepsilon)$ by considering two cases separately.

Claim OA.4. *For all i such that $V_w(T|M_i^0) < \max\limits_{y\in Y}\{y - T(y)\}$, $R_i(\varepsilon)$ is directionally differentiable at $\varepsilon = 0$ with*

$$\begin{aligned}
\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} &= -\min_{\lambda\in\Lambda_i^*(0)} \max_{F\in X_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}\right|_{\varepsilon=0} \right\} \\
&= -\max_{F\in X_i^*(0)} \min_{\lambda\in\Lambda_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon+}\right|_{\varepsilon=0} \right\}, \\
\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} &= -\max_{\lambda\in\Lambda_i^*(0)} \min_{F\in X_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon-}\right|_{\varepsilon=0} \right\} \\
&= -\min_{F\in X_i^*(0)} \max_{\lambda\in\Lambda_i^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon-}\right|_{\varepsilon=0} \right\},
\end{aligned}$$

and

$$\Lambda_i^*(0) = \left[\frac{T'(y_i-)}{1-T'(y_i-)}, \frac{T'(y_i+)}{1-T'(y_i+)}\right],$$
$$X_i^*(0) = \begin{cases} \{F \in \Delta(Y) : \mathbb{E}_F[T(y)] = T(y_i), \mathbb{E}_F[y] = y_i\}, & \text{if } y_i > \hat{y}, \\ \{F \in \Delta([0,\hat{y}]) : \mathbb{E}_F[y] \geq y_i\}, & \text{if } y_i \leq \hat{y}. \end{cases}$$

Proof. The fact that $V_w(T|M_i^0) < \max\limits_{y\in Y}\{y - T(y)\}$, implies that, for ε sufficiently small, the constraint set in (OA.11) has non-empty interior. Therefore, we can write (OA.11) as the following saddle-point problem (Luenberger, 1997)

$$R_i(\varepsilon) = -\min_{\lambda\in\mathbb{R}_+} \max_{F\in\Delta(Y)} L(F, \lambda, \varepsilon), \tag{OA.12}$$

where $L(F, \lambda, \varepsilon) = -\mathbb{E}_F[T^\varepsilon(y)] + \lambda(\mathbb{E}_F[y - T^\varepsilon(y)] - V_w(T^\varepsilon|M_i^0))$.

By Claim OA.2, the conditions in Lemma OA.2 are satisfied. Thus, $\varepsilon \to R_i(\varepsilon)$ is directionally differentiable with

$$\begin{aligned}\frac{\partial R_i(\varepsilon)}{\partial\varepsilon+}\bigg|_{\varepsilon=0} &= -\min_{\lambda\in\Lambda_i^*(0)}\max_{F\in X_i^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)]-\lambda\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial\varepsilon+}\bigg|_{\varepsilon=0}\right\}\\ &= -\max_{F\in X_i^*(0)}\min_{\lambda\in\Lambda_i^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)]-\lambda\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial\varepsilon+}\bigg|_{\varepsilon=0}\right\},\\ \frac{\partial R_i(\varepsilon)}{\partial\varepsilon-}\bigg|_{\varepsilon=0} &= -\max_{\lambda\in\Lambda_i^*(0)}\min_{F\in X_i^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)]-\lambda\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial\varepsilon-}\bigg|_{\varepsilon=0}\right\}\\ &= -\min_{F\in X_i^*(0)}\max_{\lambda\in\Lambda_i^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)]-\lambda\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial\varepsilon-}\bigg|_{\varepsilon=0}\right\},\end{aligned}$$

where $X_i^*(0)\times\Lambda_i^*(0)$ is the saddle set of $L(F,\lambda,0)$.

To complete the proof, we compute the saddle set $X_i^*(0)\times\Lambda_i^*(0)$. Let $\Lambda_i=\left[\frac{T'(y_i-)}{1-T'(y_i-)},\frac{T'(y_i+)}{1-T'(y_i+)}\right]$, and

$$X_i=\begin{cases}\{F\in\Delta(Y):\mathbb{E}_F[T(y)]=T(y_i),\mathbb{E}_F[y]=y_i\}, & \text{if } y_i>\hat{y},\\ \{F\in\Delta([0,\hat{y}]):\mathbb{E}_F[y]\geq y_i\}, & \text{if } y_i\leq\hat{y}.\end{cases}$$

Take any $(F^*,\lambda^*)\in X_i\times\Lambda_i$, and let $\tau\in[T'(y_i-),T'(y_i+)]$ be such that $\lambda^*=\tau/(1-\tau)$. Suppose first that $y_i>\hat{y}$. Then, for all $\lambda\in\mathbb{R}$,

$$L(F^*,\lambda,0)=-\mathbb{E}_{F^*}[T(y)]=L(F^*,\lambda^*,0).$$

Also, for each $F\in\Delta(Y)$ with mean $y_F\in Y$, we have

$$\begin{aligned}L(F,\lambda^*,0)=-\mathbb{E}_F[T(y)]+\frac{\tau}{1-\tau}(\mathbb{E}_F[y-T(y)]-V_w(T|M_i^0))\leq\\ -T(y_F)+\frac{\tau}{1-\tau}(y_F-T(y_F)-V_w(T|M_i^0))\leq -T(y_i)+\frac{\tau}{1-\tau}(y_i-T(y_i)-V_w(T|M_i^0))=\\ -T(y_i)=L(F^*,\lambda^*,0),\end{aligned}$$

where the first inequality follows from convexity of T, and the second inequality follows from the definition of τ which implies that $-T(y_F)+\frac{\tau}{1-\tau}(y_F-T(y_F)-V_w(T|M_i^0))$ is maximized at $y_F=y_i$. Thus, any $(F^*,\lambda^*)\in X_i\times\Lambda_i$ is a saddle point of the Lagrangian.

To show uniqueness, let $F\in\Delta(Y)$ be any solution to (OA.11) (with $\varepsilon=0$), and let $y_F\in Y$ be its mean. The fact that $y_i>\hat{y}$ implies that $y_i-T(y_i)=V_w(T|M_i^0)$: If not, by continuity of $T(y)$ and the fact that $y-T(y)$ is strictly decreasing on $(\hat{y},y_i)$, there exists $y\in(\hat{y},y_i)$ such that $y_i-T(y_i)>y-T(y)\geq V_w(M_i^0)$, contradicting the definition of y_i. We now argue that F must satisfy the constraint in (OA.11) with equality. Otherwise, $y_F-T(y_F)\geq\mathbb{E}_F[y-T(y)]>V_w(T|M_i^0)=y_i-T(y_i)$, and montonicity of $y-T(y)$ implies that $y_F>y_i$. Moreover, the fact that $y_i>\hat{y}$ implies that $T(y_F)>T(0)$. Thus, we have that $-\mathbb{E}_F[T(y)]\leq -T(y_F)<-(1-\varepsilon)T(y_F)-\varepsilon T(0)$ for $\varepsilon\in(0,1)$. Also, for ε sufficiently small, $(1-\varepsilon)(y_F-T(y_F))-\varepsilon T(0)\geq(1-\varepsilon)\mathbb{E}_F[y-T(y)]-\varepsilon T(0)\geq V_w(T|M_i^0)$. These last two statements are in contradiction with F being a solution for (OA.11).

As a result, $\mathbb{E}_F[y-T(y)]=y_i-T(y_i)$ for all $F\in X_i^*(0)$. Moreover, convexity of T implies that the optimal value of the Lagrangian is attained by a deterministic F, and therefore $\mathbb{E}_F[T(y)]=T(y_F)$ at the optimum. Combining these two facts, we have that any optimal F satisfies $y_F-T(y_F)=y_i-T(y_i)$. Further, the fact that $y_i-T(y_i)=V_w(T|M_i^0)<\overline{y}-T(\overline{y})$, together with concavity and monotonicity of $y-T(y)$, implies that $y-T(y)$ is strictly increasing a neighborhood of y_i. Thus, $y_F-T(y_F)=y_i-T(y_i)$ if and only if $y_F=y_i$. This concludes the proof that $X_i^*(0)=X_i$.

Finally, by the previous paragraph, any $\lambda\in\Lambda^*(0)$ has to be such that $y_i\in\arg\max_{y\in Y}\{-T(y)+\lambda[y-T(y)-V_w(T|M_i^0)]\}$. By convexity of T, this is achieved if and only if $\lambda\in\Lambda_i$, and thus

$\Lambda_i^*(0) = \Lambda_i$ as desired.

It remains to consider the case in which $y_i \leq \hat{y}$. By definition of $\hat{y}$, $T(y)$ is constant on $[0, \hat{y}]$, and by monotonicity and convexity of T it holds that $\min_{y \in Y} T(y) = T(\hat{y})$. Any $F \in X_i$ satisfies $\mathbb{E}_F[T(y)] = T(\hat{y})$, and

$$\mathbb{E}_F[y - T(y)] = \mathbb{E}_F[y] - T(\hat{y}) \geq y_i - T(\hat{y}) = y_i - T(y_i) \geq V_w(T|M_i^0),$$

and is therefore optimal for (OA.11). Hence, $X_i \subseteq X_i^*(0)$.

On the other hand, the definition of $\hat{y}$ implies that $T(y) > T(\hat{y})$ for all $y > \hat{y}$, and thus any F that is optimal for (OA.11) has to be supported on a subset of $[0, \hat{y}]$. Moreover, any $F \in \Delta([0, \hat{y}])$ that satisfies $\mathbb{E}_F[y] < y_i$ is not feasible—i.e., it violates $\mathbb{E}_F[y - T(y)] \geq V_w(T|M_i^0)$. These two observations imply that $X_i \supseteq X_i^*(0)$. Additionally, $y_i < \hat{y}$ implies that $\hat{y} - T(\hat{y}) > V_w(T|M_i^0)$, which implies that the constraint in (OA.11) is slack, and thus by complementary slackness, $\Lambda_i^*(0) = \{0\} = \{T'(y_i)\} = \Lambda_i$. Then, for the limiting case with $y_i = \hat{y}$, the result follows from upper hemicontinuity of the saddle set (with respect to $V_w(T|M_i^0)$). □

It remains to consider the case where $V_w(T|M_i^0) = \max_{y \in Y}\{y - T(y)\} = \overline{y} - T(\overline{y})$. In this case, condition (i) in Lemma 6 must hold and thus, for ε sufficiently small, $y - T(y) - \varepsilon D(y)$ is strictly increasing. Therefore, for all ε in a neighborhood of 0, the unique income choice maximizing type i's payoff (for any plausible M_i) is $(\delta_{\overline{y}}, 0)$. As a result, we obtain the following Claim.

Claim OA.5. *For all i such that $V_w(T|M_i^0) = \max_{y \in Y}\{y - T(y)\}$, $R_i(\varepsilon)$ is differentiable at $\varepsilon = 0$ with*

$$\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = D(\overline{y}).$$

Finally, observe that the definition of y_i implies that $y_i = \overline{y}$ and that $X_i^*(0)$, as defined in Lemma 6, is equal to $\{\delta_{\overline{y}}\}$. Applying Claim OA.5 and the above arguments, we then have that $\left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0}$ is consistent with equations (B.3) and (B.4).

Step 4: Directional derivatives of $\varepsilon \to V_P(T^\varepsilon)$. The above arguments and the chain rule can be applied to compute the directional derivative for $V_P(T^\varepsilon)$, which is given by

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0} = \int_{\mathcal{I}} \left[W_1(V_w(T|M_i^0), i) \left.\frac{\partial V_w(T^\varepsilon|M_i^0)}{\partial \varepsilon *}\right|_{\varepsilon=0} + \alpha \left.\frac{\partial R_i(\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0} \right] di, \quad * \in \{+, -\}. \tag{OA.13}$$

OA.3 Analysis of binary-types model

In the section, we formally solve for the binary-type model of Section 4.1. Let (t^*, τ^*) be the parameters of the worst-case optimal tax within the affine class as defined in Proposition 3. Let $y_0^j(\tau)$ be the income choice of type θ^j under the baseline technology and the affine tax with rate τ.

Proposition OA.1 describes a worst-case optimal tax schedule.

Proposition OA.1 (Binary types)**.** *If*

$$(\omega^2 - \alpha)(y_0^2(\tau^*) - y_0^1(\tau^*)) + \frac{\alpha \Phi(y_0^2(\tau^*), \theta^2)}{(1 - \tau^*)^2} \geq 0, \tag{OA.1}$$

then an affine tax rule defined by (t^*, τ^*) in Proposition 3 is worst-case optimal.

Otherwise, a two-piece affine tax rule is worst-case optimal. The marginal tax rates are τ^1 *for* $y \leq y_0^1$ *and* $\tau^2 > \tau^1$ *for* $y > y_0^1$*.* (τ^1, τ^2) *is given by*

$$\tau^1 = \begin{cases} 0, & \textit{if } \mathbb{E}[\omega^j] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{p^1 \alpha \Phi(y_0^1, \theta^1)}{(\alpha - \mathbb{E}[\omega^j]) y_0^1}}\right\}, & \textit{if } \mathbb{E}[\omega^j] < \alpha, \end{cases} \qquad \tau^2 = 1 - \sqrt{\frac{\alpha \Phi(y_0^2, \theta^2)}{(\alpha - \omega^2)(y_0^2 - y_0^1)}}.$$

Proof. We seek to solve the program in Lemma 5 for the case with $n = 2$. First, we find the optimal placement of the kink $\hat{y}^1 \in Y$, for a given (τ^1, τ^2). The derivative of the objective with respect to $\hat{y}^1$ is

$$\Delta(\hat{y}^1) \equiv (\tau^2 - \tau^1) \sum_{j=1}^{2} \mathbb{I}(y_0^j \geq \hat{y}^1) p^j \Big(\omega^j - \alpha + \frac{\alpha \mathbb{I}(y^j < \hat{y}^1)}{1 - \tau^1} \Big).$$

We will show that $\hat{y}^1 = y_0^1$ is optimal. If $\tau^1 = \tau^2$, then the objective is constant in $\hat{y}^1$, and therefore any choice of $\hat{y}^1$ is consistent with optimality. Next, consider the case in which $\tau^1 < \tau^2$. Without loss, we restrict attention to simple tax schedules, which implies that $\hat{y}^1 \in (y^1, y^2)$. Note that the objective is non-increasing in $\hat{y}^1$ on $[y_0^1, y^2]$, and therefore setting $\hat{y}^1 \leq y_0^1$ is optimal. Moreover, if $\Delta^1(\hat{y}^1) < 0$ and $\hat{y}^1 < y_0^1$, there is a strict welfare gain from slightly reducing $\hat{y}^1$. This follows from the fact that $\Delta^1(y)$ is left-continuous on $(0, y_0^1)$ and thus $\Delta^1(y) < 0$ in a left-neighborhood of $\hat{y}^1$. Similarly, if $\Delta^1(\hat{y}^1) > 0$ welfare strictly improves if we slightly increase $\hat{y}^1$. It remains to consider the possibility that $\Delta^1(\hat{y}^1) = 0$ and $\hat{y}^1 < y_0^1$. By inspection, $\Delta^1(y) = 0$ on $(y^1, \hat{y}^1]$ and $\Delta^1(y) < 0$ in a right-neighborhood of $y = y^1$. Thus, there is a strict welfare improvement from substituting $\hat{y}^1$ for $y^1 - \varepsilon$ where $\varepsilon > 0$ is arbitrarily small. This establishes that there exists an optimal $T(y; \hat{\mathbf{y}}, t, \boldsymbol{\tau})$ where $\hat{y}^1 = y_0^1$.

We now derive the condition under which the optimal schedule is affine. Let $\gamma \in \mathbb{R}_+$ be the Lagrange multiplier associated with the constraint $\tau^2 - \tau^1 \geq 0$. The first-order conditions are

$$-\mathbb{E}[\omega^j - \alpha] y_0^1 - p^1 \frac{\alpha \Phi(y_0^1, \theta^1)}{(1 - \tau^1)^2} - \gamma \leq 0, \quad = 0 \text{ if } \tau^1 > 0, \tag{OA.2}$$

$$-p^2(\omega^2 - \alpha)(y_0^2 - y_0^1) - p^2 \frac{\alpha \Phi(y_0^2, \theta^2)}{(1 - \tau^2)^2} + \gamma = 0, \tag{OA.3}$$

If $\omega^2 = \alpha$, then under our assumptions it must be that $\omega^1 = \omega^2 = \alpha$. (OA.2) and (OA.3) then imply that $\tau^1 = \tau^2 = 0$. Next suppose that $\omega^2 < \alpha$. We consider the cases with $\tau^1 < \tau^2$ and with $\tau^1 = \tau^2$ separately.

Case 1: $\tau^1 = \tau^2 \equiv \tau$. In this case, we have

$$\tau = \begin{cases} 0, & \text{if } \mathbb{E}[\omega^j] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{\alpha[p^1 \Phi(y_0^1, \theta^1) + p^2 \Phi(y_0^2, \theta^2)]}{p^1(\alpha - \omega^1) y_0^1 + p^2(\alpha - \omega^2) y_0^2}}\right\}, & \text{if } \mathbb{E}[\omega^j] < \alpha, \end{cases} \tag{OA.4}$$

where we have used the fact that monotonicity of ω^j and y_0^j with respect to j implies that $p^1(\alpha - \omega^1) y_0^1 + p^2(\alpha - \omega^2) y_0^2 \geq (\alpha - \mathbb{E}[\omega^j])(p^1 y_0^1 + p^2 y_0^2) > 0$ if $\alpha > \mathbb{E}[\omega^j]$, and therefore τ is well-defined.

Given this, if (and only if)

$$(\omega^2 - \alpha)(y_0^2(\tau) - y_0^1(\tau)) + \frac{\alpha \Phi(y_0^2(\tau), \theta^2)}{(1 - \tau)^2} \geq 0, \tag{OA.5}$$

then the solution to (OA.2)-(OA.3) satisfies $\gamma \geq 0$, and thus $\tau^1 = \tau^2 = \tau$. If (OA.5) does not hold, then the optimum must feature $\tau^1 < \tau^2$, which is the case that we analyze next.

Case 2: $\tau^1 < \tau^2$. In this case, $\gamma = 0$ and (OA.2) and (OA.3) give:

$$\tau^1 = \begin{cases} 0, & \text{if } \mathbb{E}[\omega^j] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{p^1 \alpha \Phi(y_0^1, \theta^1)}{(\alpha - \mathbb{E}[\omega^j]) y_0^1}}\right\}, & \text{if } \mathbb{E}[\omega^j] < \alpha, \end{cases} \tag{OA.6}$$

$$\tau^2 = 1 - \sqrt{\frac{\alpha \Phi(y_0^2, \theta^2)}{(\alpha - \omega^2)(y_0^2 - y_0^1)}}, \tag{OA.7}$$

as stated in the proposition. □

To conclude, we show that our numerical comparative statics results from Section 4.1 hold analytically under the following additional assumption which ensures that the optimal schedule from Proposition 5 is unique. First, we introduce the following assumption, which replicates those made in Mirrlees (1971) and simplifies the analysis by ensuring that the optimal taxes can be derived using first-order conditions.

Assumption 7 (Mirrleesian Φ)**.** For every $j \in \{1, 2\}$, $\Phi(\cdot, \theta^j)$ is differentiable, increasing and strictly convex. Moreover, for all $j \in \{1, 2\}$, $\lim_{y \to \overline{y}} \Phi_y(y, \theta^j) > 1$ and $\lim_{y \to 0} \Phi_y(y, \theta^j) = 0$.

Under this assumption, the necessary conditions for optimality of $(y_0^1, y_0^2, \tau^1, \tau^2, \gamma)$ are given by

$$-\mathbb{E}[\omega^j - \alpha] y_0^1 - p^1 \frac{\alpha \Phi(y_0^1, \theta^1)}{(1 - \tau^1)^2} - \gamma \leq 0, \quad = 0 \text{ if } \tau^1 > 0, \tag{OA.8}$$

$$-p^2(\omega^2 - \alpha)(y_0^2 - y_0^1) - p^2 \frac{\alpha \Phi(y_0^2, \theta^2)}{(1 - \tau^2)^2} + \gamma = 0, \tag{OA.9}$$

$$p^1 \left(\omega^1 - \alpha + \frac{\alpha}{1 - \tau^1}\right)(1 - \tau^1 - \Phi_y(y_0^1, \theta^1)) + p^2(\omega^2 - \alpha)(\tau^2 - \tau^1) = 0 \tag{OA.10}$$

$$1 - \tau^2 - \Phi_y(y_0^2, \theta^2) = 0, \tag{OA.11}$$

$$\gamma(\tau^2 - \tau^1) = 0. \tag{OA.12}$$

The marginal tax rates in a worst-case optimal simple tax are unique if the above system has a unique solution, which is what we assume next.

Assumption 8. There exists a unique $(y_0^1, y_0^2, \tau^1, \tau^2, \gamma) \in Y^2 \times [0, 1]^2 \times \mathbb{R}_+$ with $\tau^1 \leq \tau^2$ that is a solution to (OA.8)-(OA.12).

Our comparative statics results refer to the uniquely worst-case optimal (τ^1, τ^2) that arises from (OA.8)-(OA.12).

Corollary OA.1 (Comparative statics for α)**.** *Under Assumptions 7 and 8, at the worst-case optimum, τ^1 and τ^2 are non-decreasing in α (regardless of whether or not they are equal).*

Proof. Let $y_0^1(\tau^1, \tau^2; \alpha)$ and $y_0^2(\tau^2)$ be the unique solutions to (OA.10) and (OA.11). We consider two cases separately.
Case 1: (4.4) does not hold. In this region of parameter values, we have that $\tau^1 < \tau^2$ and $\gamma = 0$. If $\tau^1 = 0$, then τ^1 is (locally) constant in α. Suppose now that $\tau^1 > 0$. Applying the Implicit Function Theorem to (OA.8), we have

$$\frac{\partial \tau^1}{\partial \alpha} = \frac{y_0^1 - p^1 \Phi(y_0^1, \theta^1)/(1 - \tau^1)^2 + \frac{\partial y_0^1}{\partial \alpha}[\alpha - \mathbb{E}[\omega^j] - p^1 \alpha \Phi_y(y_0^1, \theta^1)/(1 - \tau^1)^2]}{-\{-2p^1 \alpha \Phi_y(y_0^1, \theta^1)/(1 - \tau^1)^3 + \frac{\partial y_0^1}{\partial \tau^1}[\alpha - \mathbb{E}[\omega^j] - p^1 \alpha \Phi_y(y_0^1, \theta^1)/(1 - \tau^1)^2]\}}. \tag{OA.13}$$

We begin by showing that the numerator is positive when evaluated at a solution to (OA.8)-(OA.12). To do so, we first observe that

$$\frac{\partial y_0^1}{\partial \alpha} = \frac{p^1 \frac{\tau^1}{1-\tau^1}(1-\tau^1-\Phi_y(y_0^1,\theta^1)) - p^2(\tau^2-\tau^1)}{p^1\left(\omega^1-\alpha+\frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)} =$$

$$\frac{-p^2(\omega^2\tau^1+\omega^1(1-\tau^1))(\tau^2-\tau^1)}{[\alpha\tau^1+\omega^1(1-\tau^1)]p^1\left(\omega^1-\alpha+\frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)} \leq 0,$$

where the equality follows from substituting (OA.10) and the inequality from convexity of $\Phi(\cdot,\theta^1)$. Second, suppose that $\alpha > \mathbb{E}[\omega^j]$, in which case we have

$$\alpha - \mathbb{E}[\omega^j] - \frac{p^1\alpha\Phi_y(y_0^1,\theta^1)}{(1-\tau^1)^2} \leq \alpha - \mathbb{E}[\omega^j] - \frac{\Phi_y(y_0^1,\theta^1)}{\Phi(y_0^1,\theta^1)}(\alpha - \mathbb{E}[\omega^j])y_0^1 < 0, \tag{OA.14}$$

where the first inequality follows from (OA.8), and the second one follows from strict convexity of $\Phi(\cdot,\theta^1)$, and the fact that $\Phi(0,\theta^1)=0$ and that $\alpha > \mathbb{E}[\omega^j]$ by assumption. These two facts imply that $\frac{\partial y_0^1}{\partial \alpha}[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2] \geq 0$ when $\alpha > \mathbb{E}[\omega^j]$. By an analogous argument, this result holds as well when $\alpha = \mathbb{E}[\omega^j]$. It remains to consider the first term in the numerator of (OA.13). If $\tau^1 > 0$,

$$y_0^1 - \frac{p^1\Phi(y_0^1,\theta^1)}{(1-\tau^1)^2} = \frac{E[\omega^j]y_0^1}{\alpha} \geq 0,$$

where we have substituted (OA.8). This establishes that the numerator in (OA.13) is non-negative.

Next, consider the denominator:

$$-\{-2p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^3 + \frac{\partial y_0^1}{\partial \tau^1}[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2]\} \geq$$

$$-\frac{\partial y_0^1}{\partial \tau^1}[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2] = \frac{[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2]^2}{p^1\left(\omega^1-\alpha+\frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)} \geq 0,$$

where the equality follows from differentiating (OA.10) to compute $\frac{\partial y_0^1}{\partial \tau^1}$. Therefore, $\frac{\partial \tau^1}{\partial \alpha} \geq 0$.

We now consider τ^2. By the Implicit Function Theorem,

$$\frac{\partial \tau^2}{\partial \alpha} = \frac{y_0^2 - y_0^1 - \Phi(y_0^2,\theta^2)/(1-\tau^2)^2 + \frac{\partial y_0^1}{\partial \alpha}(\omega^2-\alpha)}{-\{-2\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^3 + \frac{\partial y_0^1}{\partial \tau^2}(\omega^2-\alpha) + \frac{\partial y_0^2}{\partial \tau^2}[\alpha-\omega^2-\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^2]\}}. \tag{OA.15}$$

Considering the numerator first, we have

$$y_0^2 - y_0^1 - \Phi(y_0^2,\theta^2)/(1-\tau^2)^2 + \frac{\partial y_0^1}{\partial \alpha}(\omega^2-\alpha) \geq y_0^2 - y_0^1 - \Phi(y_0^2,\theta^2)/(1-\tau^2)^2 = \frac{\omega^2(y_0^2-y_0^1)}{\alpha} \geq 0,$$

where the first inequality follows from the fact that $\frac{\partial y_0^1}{\partial \alpha}$ and $\alpha > \omega^2$, and the equality follows from substituting (OA.9). Hence, the numerator is positive.

Similarly, for the denominator of (OA.15), we can write

$$\begin{aligned}
-\{-2\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^3 + \frac{\partial y_0^1}{\partial \tau^2}(\omega^2-\alpha) + \frac{\partial y_0^2}{\partial \tau^2}[\alpha-\omega^2-\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^2]\} \geq \\
-\frac{\partial y_0^1}{\partial \tau^2}(\omega^2-\alpha) - \frac{\partial y_0^2}{\partial \tau^2}[\alpha-\omega^2-\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^2] = \\
\frac{p^2(\omega^2-\alpha)^2}{p^1\left(\omega^1-\alpha+\frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)} - \frac{\alpha-\omega^2-\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^2}{\Phi_{yy}(y_0^2,\theta^2)} = \\
\frac{p^2(\omega^2-\alpha)^2}{p^1\left(\omega^1-\alpha+\frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)} - \frac{-\omega^2-\alpha\tau^2/(1-\tau^2)}{\Phi_{yy}(y_0^2,\theta^2)} > 0,
\end{aligned}$$

where the first equality follows from substituting the expression for $\frac{\partial y_0^1}{\partial \tau^2}$ and $\frac{\partial y_0^2}{\partial \tau^2}$, and the second equality follows from substituting in (OA.11). We have thus established that $\frac{\partial \tau^j}{\partial \alpha} \geq 0$ for all $j = 1, 2$.
Case 2: (4.4) holds with strict inequality. Here, $\tau^1 = \tau^2 \equiv \tau^*$ and the planner's objective is

$$V(\tau) = \sum_{j=1}^{2} p^j \left(\omega^j + \frac{\alpha\tau}{1-\tau}\right)((1-\tau)y_0^j(\tau) - \Phi(y_0^j(\tau),\theta^j)),$$

where $y_0^j(\tau)$ is defined implicitly by $1-\tau = \Phi_y(y_0^j(\tau),\theta^j)$. The optimal τ^* is given by (OA.4). By continuity, it suffices to show that τ^* is increasing in the region in which $\tau^* > 0$. The cross partial derivative of $V(\tau)$ with respect to (τ,α) at τ^* is

$$\frac{\partial V'(\tau^*)}{\partial \alpha} = \sum_{j=1}^{2} p^j \left(y_0^j(\tau^*) - \frac{\Phi(y_0^j(\tau^*),\theta^j)}{(1-\tau^*)^2}\right) = \sum_{j=1}^{2} p^j \frac{\omega^j y_0^j(\tau^*)}{\alpha} > 0,$$

where the second equality follows from plugging in (OA.4). Therefore, applying the univariate Implicit Function Theorem, we obtain that τ^* is non-decreasing in α in the region where $\tau^1 = \tau^2 = \tau^* > 0$.

By Berge's Maximum Theorem, the (unique) maximizer is continuous in α, and therefore monotonicity of (τ^1,τ^2) continues to hold at the point in which (4.4) holds with equality. □

Corollary OA.2 (Comparative statics for ω^2)**.** *Under Assumptions 7 and 8, at the worst-case optimum, τ^1 and τ^2 are non-increasing in ω^2 (regardless of whether or not they are equal). Moreover, there exists a cutoff $\omega \in (0,\alpha)$ such that if $\mathbb{E}[\omega^j] \in (\omega,\alpha]$, then $\tau^2 - \tau^1$ is non-increasing in ω^2.*

Proof. Case 1: (4.4) does not hold. If $\tau^1 = 0$, then it is (locally) constant in ω^2. Suppose now that $\tau^1 > 0$. Similar to the proof of Corollary OA.1, we have

$$\frac{\partial \tau^1}{\partial \omega^2} = \frac{-p^2\omega^2 y_0^1 + \frac{\partial y_0^1}{\partial \omega^2}[\alpha-\mathbb{E}[\omega^j]-p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2]}{-\{-2p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^3 + \frac{\partial y_0^1}{\partial \tau^1}[\alpha-\mathbb{E}[\omega^j]-p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2]\}}. \quad \text{(OA.16)}$$

We already established in the proof of Corollary OA.1 that the denominator of (OA.16) is positive.

For the numerator, we have

$$-p^2\omega^2 y_0^1 + \frac{\partial y_0^1}{\partial \omega^2}[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2] =$$

$$-p^2\omega^2 y_0^1 + \frac{p^2(\tau^2-\tau^1)}{p^1\left(\omega^1 - \alpha + \frac{\alpha}{1-\tau^1}\right)\Phi_{yy}(y_0^1,\theta^1)}[\alpha - \mathbb{E}[\omega^j] - p^1\alpha\Phi_y(y_0^1,\theta^1)/(1-\tau^1)^2] < 0,$$

where the equality follows from computing $\frac{\partial y_0^1}{\partial \omega^2}$ using the Implicit Function Theorem, and the inequality follows from (OA.14). Hence, $\frac{\partial \tau^1}{\partial \omega^2} < 0$.

Additionally,

$$\frac{\partial \tau^2}{\partial \omega^2} = \frac{-1 - \frac{\partial y_0^1}{\partial \omega^2}(\alpha - \omega^2)}{-\{-2\alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^3 + \frac{\partial y_0^1}{\partial \tau^2}(\omega^2-\alpha) + \frac{\partial y_0^2}{\partial \tau^2}[\alpha - \omega^2 - \alpha\Phi_y(y_0^2,\theta^2)/(1-\tau^2)^2]\}} < 0. \tag{OA.17}$$

Case 2: (4.4) holds with strict inequality. In this case, using the notation from the proof of Corollary OA.1, we have

$$\frac{\partial V'(\tau^*)}{\partial \omega^2} = -p^2 y_0^2(\tau^*) < 0,$$

and therefore τ^* is decreasing in ω^2 in this region as well. Monotonicity at the point in which (4.4) holds with equality follows again from continuity of the maximizer. This establishes the first part of the Corollary involving the monotonicity of τ^1 and τ^2.

Finally, to show the second part of the result, consider the region of parameters such that $\mathbb{E}[\omega^j]$ is arbitrarily close to α.[45] In that region, we have that $\tau^1 = 0$ and τ^2 is weakly decreasing in ω^2 (strictly so unless $\tau^2 = \tau^1 = 0$). Therefore, it holds that $\tau^2 - \tau^1$ decreases with ω^2 in that region. □

[45] To ensure existence of the optimal tax, we continue to restrict attention to the case in which $\mathbb{E}[\omega^j] \leq \alpha$.